PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 12 June 15, 1926 Number 6

BENDING OF A CURVED TUBE OF CIRCULAR CROSS SECTION[1]

By William Hovgaard

Massachusetts Institute of Technology

Communicated April 30, 1926

It has long been known that pipe bends are much more flexible than can be explained by the ordinary formula for bending of curved bars. Experiments by Professor A. Bantlin of Stuttgart[2] show that the displacement of the ends of double-offset pipe-bends, when subject to forces tending to alter the curvature, may be as much as five times as great as found by the ordinary theory. Professor Bantlin ascribed this to corrugations on the concave side and stretching on the convex side produced in manufacture, but Dr. Th. v. Kármán of Göttingen advanced the theory that the principal cause of the discrepancy was a deformation of the pipe by which the tube section under bending takes a slightly flattened, oval form.

It can be seen at once that when a pipe is subject to a bending couple, tending for instance to increase the curvature, the stress forces will have a tendency to push the parts on both sides of a transverse section towards the neutral surface. Hereby the material farthest from the neutral axis is enabled to shirk its duty and is largely relieved of stresses, while the maximum stresses will occur at points much nearer the neutral axis. On this theory Dr. v. Kármán developed formulas by which the augmented deflections could be largely explained.

Professor H. Lorenz of Danzig[3] offered a quite different solution based on the theory of thin shells and arrived at the conclusion that the stresses directly due to bending of the pipe follow the usual linear law. Deflections calculated according to this method showed a good agreement with Bantlin's experiments.

Other investigators[4] have used the ordinary formula for bending of curved bars and have made experiments which seemed to show a fair agreement with the calculated deflections. It must be noted, however, that in these experiments the end flanges were subject to restraint, whereby unknown couples were liable to be induced in the pipe bend.

In order to determine which of the above-mentioned theories was cor-

rect, the author of this paper worked out a solution by a new method, as explained below, and applied the resulting formulas to tests on pipe bends carried out under his supervision at the Massachusetts Institute of Technology. In these tests were measured not only the deflections of the ends of the bends, as in former tests, but also the strains (stresses) and deformations at certain sections. As in Bantlin's tests the end flanges were free to turn about knife edges placed normal to the plane of the bend. The measurements confirmed fully the theory; the deformations, displacements and stresses being in close agreement with the calculated values. The formulas were also confirmed by an analysis of Bantlin's experiments. Figure 1 shows the calculated stress distribution over one-half of the section of a 200-mm. pipe and for comparison the ordinary straight-line stress diagram is shown on the same figure.

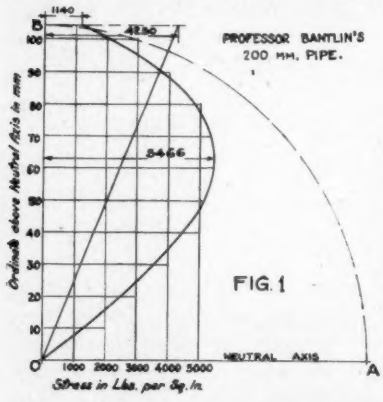


FIG. 1

Consider an angular element, of angle $d\psi$, of a curved pipe bent to a radius R, of circular section with radius r and moment of inertia I, subject to a bending moment M. Let E be the modulus of elasticity. Then, according to the ordinary formula, the angular flexural strain ratio is given by

$$\frac{\Delta d\psi}{d\psi} = \frac{R\,M}{I\,E} \tag{1}$$

while, as shown in the Appendix, we obtain by the new method:

$$\frac{\Delta d\psi}{d\psi} = \frac{R\,M}{I\,E}\,\frac{48h^2R^2 + 10r^4}{48h^2R^2 + r^4} \tag{2}$$

where h is one-half the wall thickness, r is the radius to the middle surface of the wall.

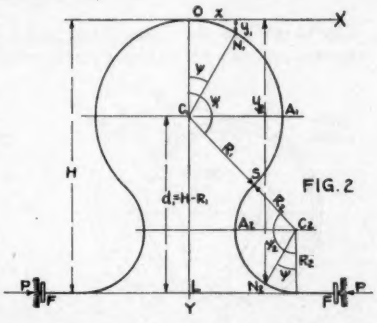
The longitudinal direct fiber stress is found to be:

$$p = \frac{E}{R}\,\frac{d\,\Delta\psi}{d\psi}\left(y - \frac{6r^2}{24h^2R^2 + 5r^4}\,y^3\right) \tag{3}$$

which is seen to differ from the ordinary formula for bending stresses in the addition of the second term inside parentheses.

Formulas (2) and (3) were obtained after having discarded terms of

relatively small value and were then found to be of the same form as those of Dr. v. Kármán. Being derived by an entirely different method and corroborated by experiments they are believed to represent a correct approximate solution.



In calculating the deflection of a pipe bend of the ordinary U-type or of the double-offset type (Fig. 2), when subject to axial forces at the end flanges, we place a system of rectangular coördinates with origin at the top of the bend, OX tangent to the axis and OY along the line of symmetry of the bend. Let the entire height of the bend be H and suppose a force P is acting on each end flange parallel to OX tending to compress the bend. Then the bending moment at any point is:

$$M = P(H - y)$$

and the displacement of one of the end flanges, F, relative to OY is:

$$\Delta x = \int_0^F (H - y)d(d\psi) = \frac{P}{EI}\int_0^F R(H - y)^2 \frac{48h^2R^2 + 10r^4}{48h^2R^2 + r^4}\, d\psi. \quad (4)$$

Where the bend is made up of fairly regular circular arcs, this integral may be evaluated algebraically, but if the bends are not true circular arcs it is better to find the value of the integral graphically. In the analysis made by the author of Professor Bantlin's tests and of the experiments made at the Massachusetts Institute of Technology both methods were used.

APPENDIX: DEFORMATION OF A SECTIONAL ELEMENT OF A CURVED TUBE SUBJECT TO BENDING

Let R be the radius of curvature of the axis of the tube in the unstrained state, r the radius of the middle surface of the tube wall, and $2h$ the thickness of the wall. We assume that $2h$ is small relative to r and that r is small relative to R. The axis of the tube as well as the middle surface of the tube wall are regarded as inextensible. The tube is not subject to any pressure, external or internal.

The tube is conceived as a body of revolution formed by revolving a transverse section about an axis $X'X'$ as indicated in figure 3. Consider a segment of the tube enclosed between two planes through $X'X'$, forming an angle $d\psi$ with each other. OO' is an element of the axis of the tube and $PQQ'P'$ is an element of the tube wall, having sectional end areas $2hrd\theta$. Let p be the longitudinal stress created by bending, then a force $2phrd\theta$ will be acting at each end of the element forming an angle $d\psi$ with each other. The resultant of these forces, f, acts everywhere normal to the axis of revolution and hence normal to the neutral axis of the tube. If the bending couple is such as to

increase the curvature, the forces are all directed towards the neutral surface. We have

$$f = 2hrpd\theta d\psi. \tag{1}$$

Due to bending, the radius of curvature R is changed to R', $d\psi$ is changed to $d\psi + \Delta d\psi$; the ordinate y of a point P on the middle surface becomes $y - \Delta y$ due to the flattening of the section. We find the elongation per unit length of the filament PP', approximately,

$$\epsilon = \frac{y}{R}\frac{\Delta d\psi}{d\psi} - \frac{\Delta y}{R}. \tag{2}$$

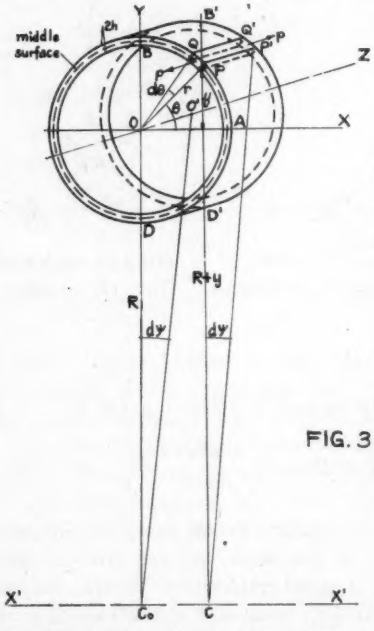


FIG. 3

It is important to note that f always acts normal to the neutral surface; it produces compression, shearing and bending stresses in the tube wall, but these stresses will not appreciably affect p.

We may therefore write

$$p = E\epsilon = \frac{E}{R}\left(y\,\frac{\Delta d\psi}{d\psi} - \Delta y\right). \tag{3}$$

Making the elementary filament of unit length we write $d\psi = \dfrac{1}{R+y}$ or approximately $\dfrac{1}{R}$, whence substituting in (1) from (3) we get

$$f = \frac{2\,hr\,E}{R^2}\left(y\,\frac{\Delta d\psi}{d\psi} - \Delta y\right)d\theta. \tag{4}$$

The further development depends upon making an assumption as to the value of Δy. It was first tentatively assumed that $\Delta y = ky$ where k is a constant over the section, but this led to the inconsistent result: $\Delta y = ky^3$. Hence the latter value for Δy was adopted and led as shown below to an expression for Δy which contained also a term to the fifth power of y but did not differ much in value from ky^3. We assume then

$$\Delta y = ky^3 = kr^3 \sin^3\theta \tag{5}$$

$$\therefore f = \frac{2hr^2E}{R^2}\left(\frac{\Delta d\psi}{d\psi} - kr^2\sin^2\theta\right)\sin\theta. \tag{6}$$

We now by usual methods calculate the deflection Δy of a transverse element of the tube due to the bending moment m exerted by the f-forces. By symmetry we need only consider one quadrant. The conditions that the sum of the angular deflections over the quadrant must be zero by symmetry and that the sum of the f-forces must equal the reaction at the neutral surface, furnish the unknown reactions. We find:

$$m = \frac{hr^3E}{2R^2}\left[\frac{\Delta d\psi}{d\psi}(1-2\cos^2\theta) - \frac{1}{3}\,kr^2\left(\cos^4\theta - 6\cos^2\theta + \frac{21}{8}\right)\right]. \tag{7}$$

The vertical deflection at any point P is found from:

$$\Delta y = -\int_0^{\theta_p} \frac{m}{iE}(x - x_p)\, ds \tag{8}$$

where $i = \frac{1}{3}h^3$ is the moment of inertia of the section of the tube wall.

$$\therefore \Delta y = \frac{r^4}{4h^2R^2}\left[\frac{\Delta d\psi}{d\psi}\sin^3\theta - kr^2\left(\frac{19}{24}\sin^3\theta + \frac{1}{20}\sin^5\theta\right)\right]. \tag{9}$$

We have

$$k = \frac{\Delta y_B}{r^3} \tag{10}$$

and for $\theta = \frac{\pi}{2}$ we find from (9):

$$\Delta y_B = \frac{r^5}{4h^2R^2}\left(\frac{\Delta d\psi}{d\psi} - \frac{101}{120}\frac{\Delta y_B}{r}\right)$$

which on solving for Δy_B, gives:

$$\Delta y_B = \frac{r^5}{4h^2R^2 + \frac{101}{120}r^4}\frac{\Delta d\psi}{d\psi}. \tag{11}$$

Substituting in (9)

$$\Delta y = \frac{r^5\sin^3\theta\left(4h^2R^2 + \frac{1}{20}r^4\cos^3\theta\right)}{4h^2R^2\left(4h^2R^2 + \frac{101}{120}r^4\right)}\frac{\Delta d\psi}{d\psi}.$$

Simplifying by the omission of small terms we may write:

$$\Delta y = \frac{6r^5\sin^3\theta}{24h^2R^2 + 5r^4}\frac{\Delta d\psi}{d\psi} \tag{12}$$

which is consistent with the assumption made in (5). Substituting in (3) from (12) we find the stress:

$$p = \frac{E}{R}\left(y - \frac{6r^2y^3}{24h^2R^2 + 5r^4}\right)\frac{\Delta d\psi}{d\psi}. \tag{13}$$

In order to determine the angular deflection we equate the external bending moment to the moment about the neutral axis of the stress forces:

$$M = \int_0^{2\pi} 2hrp\,(y - \Delta y)\, d\theta$$

and neglecting terms in $\left(\frac{\Delta d\psi}{d\psi}\right)^2$ we find on solving for the angular deflection:

$$\frac{\Delta d\psi}{d\psi} = \frac{MR}{EI}\frac{48h^2R^2 + 10r^4}{48h^2R^2 + r^4}. \tag{14}$$

We are now able to find the stress at any point from (13) and the displacement of the ends of a pipe bend from formula (4) in the main body of the paper.

[1] A more detailed account will appear in *J. Math. Physic.*, Mass. Inst. Tech., Vol. 6, No. 1 (1926).

[2] *Mittheilungen über Forschungsarbeiten*, Heft 96, 1910.

[3] Dingler's *Polyt. Journal*, **1912.**

[4] S. Crocker and S. S. Sanford, *Amer. Soc. Mech. Eng.*, **1922.**

THE HEAT CAPACITY OF SOLID ALIPHATIC CRYSTALS. II

BY E. O. SALANT*

JOHNS HOPKINS UNIVERSITY

Communicated May 10, 1926

In the preceding issue of these PROCEEDINGS, the fact of characteristic frequencies for non-polar bonds was introduced into the Born-Karman heat capacity equations, which were then, with certain approximations, brought into a general simplified form for non-polar compounds.[1] With the aid of these simplified equations, the heat capacities of some organic solids will now be calculated and compared with their observed heat capacities.

Due to the still scanty knowledge of the infra-red vibrations of organic substances, it will be necessary to make use of approximations to get the required frequencies, and so agreements and disagreements of the calculations with the measurements cannot be taken as severe tests of the validity of the equations.

For the molecular heat capacity at constant volume, C_v, of an open-chain compound $A_xB_yC_z$ having n atoms and b non-polar bonds to the chemical molecule, we had

$$C_v = f'_D(\nu_m/T) + 2[(x-1)E'(\nu^{(A)}/T) + yE'(\nu^{(B)}/T) + zE'(\nu^{(C)}/T)] + \sum_{j=1}^{b} E'(\nu_j^{(b)}/T) \quad (1)$$

where $f'_D(\nu_m/T)$ is the Debye atomic heat capacity equation and $E'(\nu/T)$ is the Einstein function for a linearly oscillating atom. These are defined as

$$\left.\begin{aligned} f'_D(\nu_m/T) &= \frac{9R}{(\theta_m/T)^3}\int_0^{\theta}\frac{(\theta/T)^4 e^{\theta/T}\,d\theta}{(e^{\theta/T}-1)^2} \\ E'(\nu/T) &= \frac{R(\theta/T)^2 e^{\theta/T}}{(e^{\theta/T}-1)^2};\ \theta = \frac{h\nu}{k} \end{aligned}\right\} \quad (2)$$

where T is the temperature, h is Planck's constant, R the absolute gas constant and k the constant per molecule (R divided by the Avogadro number).

Equation (1) will now be limited to the case of aliphatic compounds of C, H and O. It is necessary to choose first some atom as nearest to the center of the molecule, that is, some atomic species corresponding to the species A. For the substances considered in this paper, a C atom will be chosen; this may not always be proper (for example, if methyl or ethyl ether were to be considered, the O atom would have to be the central atom). We have, then, for the compound $C_xH_yO_z$

$$C_v = f'_D(\nu_m/T) + 2[(x-1)E'(\nu^{(C)}/T) + yE'(\nu^{(H)}/T) + zE'(\nu^{(O)}/T)] + \sum_{j=1}^{b} E'(\nu_j^{(b)}/T) \quad (3)$$

where ν_m refers to the molecular vibration $\nu^{(C)}$, $\nu^{(H)}$ and $\nu^{(O)}$ to the vibrations of the C, H and O atoms across the bonds, and the last term to the bond vibrations. Equations (2) and (3) permit the calculation of the heat capacities of aliphatic solids if the required frequencies are known; methods of obtaining values for these frequencies will now be discussed.

In the first attempt to apply the quantum equations of heat capacities of solids to aliphatic compounds, the writer found it necessary to use a mean vibration in all directions for the atoms and to obtain this frequency with the Koref-Lindemann melting-point formula.[2] If the atoms are thought of, first as vibrating with the frequencies given by this formula, but then restricted to the characteristic bond frequencies along the lines of the bonds, and the effect of the bond on the other vibrations is neglected, then this formula may be used for calculating the vibrations of the atoms across the bonds. Since there is no method as yet for obtaining those frequencies (infra-red spectra being still inadequate), the Koref-Lindemann formula had to be used in this way for the $2(x + y + z)$ frequencies. The writer regards this as the greatest source of error in these calculations.

The mean frequency ν_m for the Debye equation is determined in the strict theory by the elastic properties of the crystal, but measurements of elastic constants of organic compounds are, unfortunately, lacking. For those compounds whose heat capacities have been measured at the lowest temperatures (glucose, for example, at 20°K.), where the Einstein functions become negligible, ν_m can be obtained from the heat capacity measured at some one of these lowest temperatures. Again, unfortunately, too few organic heat capacity measurements have been carried to these low temperatures.

The remaining alternative is to use the melting-point equation, following Nernst,[3] with the molecular weight M in place of the usual atomic weight:

$$\nu_m = 3.1 \times 10^{12} T_S^{1/3} d^{1/3} M^{-1/4}$$

where T_S is the absolute melting-point and d is the density of the compound.

For our purposes, this may be approximated still further. The density of organic compounds being usually not far from 1, we may write

$$\nu_m = 3.1 \times 10^{12} T_S^{1/2} M^{-1/6}$$

which enables us to evaluate the Debye equation for all those organic compounds whose molecular weight and melting-point are known. For comparison, we have the molecular heat at a very low temperature of glucose only: at $T = 20$, from the melting-point formula and Debye equation, $\bar{c} = 0.18$, measured 0.09.

The characteristic frequencies of the C—H, C—C, C—O and O—H bonds are required for our calculations. The small variations of these frequencies with variations in the degree of saturation and loading of the atoms of a bond will be neglected.

The absorption band of the C—H bond was found at 6.9μ, of the C—C at 28.0μ, both by Ellis.[4]

The absorption of the alcoholic OH group was found at 2.95μ by Coblentz,[5] at 3.0μ by Weniger.[6] Coblentz further demonstrated that this band disappears in organic acids, and neither in his nor in Weniger's measurements of the acids can any band characteristic of carboxylic OH be detected. The band at 3μ will, therefore, be taken as the fundamental for the alcoholic O—H bond.

Coblentz[7] lists a band at 29.4μ as characteristic of carbonates. The measurements of Coblentz and of Schaefer and Shubert[8] showed a band varying from 13.32 to 15.20μ for various carbonates, and this may be taken as the approximate second harmonic of the 29μ band as a fundamental (first harmonic).

If we assume that these are due to the C—O bond, evidence should appear in the spectrum of CO_2. Rubens and Aschkinass[9] found an emission band of CO_2 at 14.1μ, and absorption at 14.7μ; Burmeister[10] showed the absorption to be a double band, 14.70 and 15.05μ.

Weniger attributed an absorption band at 5.9μ to the C=O group, and Marsh[11] gives further evidence of a C=O vibration at 5.8μ. It is to be noted that this would be a fifth harmonic of a fundamental 29μ C—O band.

Weniger found a band from 9.1 to 9.8μ, common to alcohols, esters and organic acids, which he could not classify (this being before the assignment of bands to bonds, initiated by Ellis). The one bond common to all of these (in addition to the C—H with band at 6.9μ) is the C—O, and a band in this region is, therefore, to be looked for in ethers. Coblentz's curves[12] show absorption bands at about 9.3μ for both methyl and ethyl ether vapors. The band at 9.1 to 9.8μ characteristic of the C—O bond would be an approximate third harmonic of the fundamental 29μ.

Thus this fundamental serves to correlate bands in CO_2, carbonates,

ketones, ethers, alcohols, acids and esters, and will, therefore, be taken as the characteristic of the C—O bond.

The $h\nu/k$ values of these bonds are:

C—H	C—C	C—O	O—H
2115	521	503	4863

It is to be noted that the contributions to the heat capacity of the O—H bond are inappreciable below 347°K. For the carboxyl O—H, in formic and oxalic acids, the H atoms are given a mean vibration in all directions, calculated by the Koref-Lindemann formula.

Thus the mean atomic heat at constant volume, $\bar{c}_v(=C_v/n)$, was obtained by equations (2) and (3), the melting-points, molecular weights and

GLYCEROL

T	$\bar{c}_p$ CALC.	OBS.[14]
280	3.09	2.46
227	2.54	1.79
174	1.97	1.44
116	1.25	1.07
70	0.69	0.72

OXALIC ACID

T	$\bar{c}_p$ CALC.	OBS.[15]
200	3.13	2.30
160	2.63	1.95
120	2.00	1.60
100	1.60	1.41

GLUCOSE

T	$\bar{c}_p$ CALC.	OBS.[17]
300	3.06	2.29
260	2.69	1.95
220	2.34	1.64
180	1.93	1.34
140	1.40	1.06
100	0.84	0.78
60	0.34	0.49
20	0.18	0.09

FORMIC ACID

T	$\bar{c}_p$ CALC.	OBS.[15]
270	3.60	3.42
237	3.39	3.02
205	3.18	2.73
176	2.97	2.48
94	2.06	1.73
71	1.57	1.45

METHYL ALCOHOL

T	$\bar{c}_p$ CALC.	OBS.[13]
152	2.09	2.36
112	1.64	1.91
101	1.53	1.77
89	1.27	1.61

ETHYL ALCOHOL

T	$\bar{c}_p$ CALC.	OBS.[16]
140	1.78	1.81
130	1.64	1.60
120	1.51	1.51
110	1.39	1.42
100	1.26	1.31
90	1.15	1.25

BUTYL ALCOHOL

T	$\bar{c}_p$ CALC.	OBS.[13]
151	1.63	1.35
145	1.52	1.32
111	1.10	1.11
100	0.96	1.03
91	0.85	0.95

ACETONE

T	$\bar{c}_p$ CALC.	OBS.[18]
173	2.32	2.32
153	2.05	2.09
133	1.74	1.80
113	1.45	1.64
93	1.17	1.22

characteristic bond frequencies. The mean atomic heat at constant pressure, $\bar{c}_p$, was then calculated by the Nernst-Lindemann relation:[19]

$$c_p - c_v = c_p^2 T A_o / T_S$$

where $A_o = 0.0214$. This, again, had to be used in place of the thermodynamic relation, due to the lack of measurements of thermal expansions and compressibilities of organic solids.

Tables of the calculated and observed heat capacities of methyl, ethyl and butyl alcohols, glycerol, acetone, glucose and formic and oxalic acids are given, from the lowest to the highest temperatures measured.

Comparison of the results of this paper with those of the preceding one on aliphatic heat capacities[2] shows that the calculated values of the three alcohols, of acetone and of formic acid are now uniformly nearer the experimental values, the improvement at lower temperatures being especially marked. At these temperatures, the calculated heat capacities of glycerol, glucose and oxalic acid are also better, but at higher temperatures are now worse. However, in the calculations of that paper a special assumption as to the C vibrations in these three substances was made; without this special assumption, their values even at higher temperatures are now improved (for example, error in glucose at 300°K. formerly 50%, now 30).

It may be concluded, then, that by introducing asymmetric vibrations for the atoms due to their bonds, and also molecular vibrations, the calculated heat capacities of these organic compounds show consistently better agreement with measurement. These and other results are to be more fully discussed at a later date.

* Johnston Scholar.

[1] E. O. Salant, *Proc. Nat. Acad. Sci.*, **12** (1926).
[2] E. O. Salant, *Ibid.*, **11**, 227 (1925).
[3] W. Nernst, *Theory of the Solid State*, Univ. of London Press, **1924**, §12.
[4] J. W. Ellis, *Phys. Rev.*, **23**, 48 (1924); **27**, 298 (1926).
[5] W. W. Coblentz, *Carnegie Institute of Washington Pub.*, No. **35** (1905), p. 117.
[6] W. Weniger, *Phys. Rev.*, **31**, 388 (1910).
[7] W. W. Coblentz, *Carnegie Institute of Washington Pub.*, No. **97** (1905), p. 67.
[8] C. Schaefer and M. Shubert, *Ann. Physik.*, **50**, 283 (1916).
[9] H. Rubens and E. Aschkinass, *Astrophys. J.*, **8**, 191 (1898).
[10] W. Burmeister, *Ber. deutsch physik. Ges.*, **11**, 589 (1913).
[11] J. W. Marsh, *Phil. Mag.*, **294**, 1206 (1925).
[12] W. W. Coblentz, *Carnegie Pub.*, No. **35**, figures 17 and 18.
[13] G. S. Parks, *J. Am. Chem. Soc.*, **47**, 338 (1925).
[14] Gibson and Glauque, *Ibid.*, **93**, 45 (1923).
[15] Gibson, Latimer and Parks, *Ibid.*, **42**, 1533 (1920).
[16] O. Maass and J. Waldbauer, *Ibid.*, **47**, 1 (1925).
[17] F. Simon, *Ann. Physik*, **68**, 241 (1923).
[18] W. Nernst, *Ibid.*, **36**, 395 (1911).
[19] W. Nernst and F. A. Lindemann, *Zs. f. Elektrochem.*, **17**, 817 (1911).

THE EFFECT OF CRYSTAL SIZE UPON THE INTENSITY OF X-RAY REFLECTION

BY R. J. HAVIGHURST[1]

CRUFT LABORATORY, HARVARD UNIVERSITY

Communicated May 3, 1926

It appears possible to determine with considerable accuracy the electron distribution in atoms[2] from a knowledge of the intensity of reflection of X-rays by crystals containing those atoms. However, the effect of certain other factors, notably the size of the reflecting crystal, upon the intensity of reflection must be known before conclusions can be drawn concerning electron distribution.

Darwin[3] has studied theoretically the effect of crystal size upon the intensity of X-ray reflection. His expression for the intensity of reflection, I, at an angle θ from the face of a *small perfect* crystal may be written

$$I \propto N^2 f^2 \lambda^3 \frac{(1+\cos^2 2\theta)}{\sin 2\theta} e^{-\frac{b^2 \sin^2\theta}{\lambda^2}} = A^2 \frac{(1+\cos^2 2\theta)}{\sin 2\theta} = Q. \qquad (1)$$

N is the number of atoms per unit volume; f is a function of $\frac{\lambda}{\sin\theta}$ and of the electron distribution in the atoms; $e^{-\frac{b^2 \sin^2\theta}{\lambda^2}}$ is the Debye temperature factor. If the crystal has a linear dimension larger than 10^{-5} cm., the intensity of reflection is modified by "extinction." Before considering extinction, let us define a *perfect crystal* as one in which the planes of each unit cell are perfectly parallel to the corresponding planes of every other unit cell, so that the whole crystal can be regarded as a macroscopic copy of the unit cell. Darwin distinguishes two kinds of extinction, *primary* extinction, which occurs within each little block that acts like a perfect crystal, and *secondary* extinction, which represents the shielding of the blocks deep within the crystal by the reflection of the X-rays from the blocks near the surface. The effect of primary extinction is represented by the correction factor in the expression

$$Q' = Q \frac{\tanh \sqrt{2Q \frac{d^2 \cot\theta}{\lambda}}}{\sqrt{2Q \frac{d^2 \cot\theta}{\lambda}}}. \qquad (2)$$

Q' is proportional to the *measured* intensity of reflection and becomes equal to Q in (1) when extinction is negligible; d is the thickness of the crystal. If expression (2) is evaluated for rock-salt, using the value of Q', obtained by Bragg, James and Bosanquet for the (200) reflection[4]

and corrected for secondary extinction, as the Q under the radical, it is found that for $d = 5 \times 10^{-5}$ cm., the correction factor is 0.96, while for $d = 5 \times 10^{-3}$ cm., the factor is 0.03. Since the value of Q used to obtain these figures is really a Q' value and may have been reduced by primary extinction, it is evident that these are maximum values for the correction factor.

Secondary extinction is appreciable only in crystals larger than 10^{-2}–10^{-3} cm. and is directly proportional to the intensity of reflection. Bragg, James and Bosanquet have been able to correct satisfactorily for its effect in their measurements upon large single crystals of rock-salt. Darwin suggested that information regarding extinction might be obtained by measuring the intensity of reflection from crystal powders of varying sizes. If each powder particle is a perfect crystal, the intensity of reflection would vary with the size until the particles were less than 5×10^{-5} cm. in thickness. Further, the relative intensities for different planes would vary, because the extinction depends upon the intensity and angle of reflection.

Evidence which has been gathering from different sources[5] indicates that in spite of its flawless appearance, a macroscopic crystal is almost never perfect, but is a *mosaic* structure consisting of many very small perfect crystals with very nearly but not exactly the same orientation. The size of these perfect crystal blocks and their divergence from the mean orientation depend upon the composition of the crystal and upon the mechanical and thermal treatment which it has undergone. Mark estimates the linear dimensions of the perfect crystal blocks in most crystals to be about 10^{-4} cm. Smekal estimates the thickness of the perfect crystals in rock-salt to be 4×10^{-5} cm. Becker obtains a value of less than 10^{-5} cm. for tungsten.

That there is some primary extinction operating to reduce the intensity of reflection from a large single crystal is evidenced by the fact that grinding a crystal face increases the intensity of reflection to a marked degree. Mechanical strains probably break up the crystal blocks near the surface or produce distortions in them which reduce primary extinction. Grinding a crystal to a powder, then, might be expected to distort the powder particles so that the size of the perfect crystals would be many times smaller than the size of the particles themselves. The experiments described below were carried out in an attempt to decide whether the reflection of X-rays by powdered crystals is free from the effects of extinction.

Experimental.—The relative intensities of reflection of the Mo Kα doublet by the different planes of many crystal powders were determined, using the radiation, filtered through enough ZrO_2 to reduce Mo K$(\gamma + \beta)$ to 0.4% of Mo K$(\alpha_1 + \alpha_2)$, from a Mo tube operated at a constant potential of 35,000 volts.[6] The primary beam was reflected from the face of a briquet of the powder under investigation into an ionization chamber, the distances from focal spot to powder, a, and from powder to ionization

chamber slit, b, determining the ratio $\frac{\sin \alpha}{\sin \beta} = \frac{a}{b}$ where α is the angle between the primary beam and the surface of the powder and $\beta = 2\theta - \alpha$. This arrangement, which has been described fully by Brentano,[7] causes a sharp focussing of the reflected beam, even though a primary beam of wide divergence is used. With this apparatus, it is possible to measure the intensities of reflection out to $\theta = 45°$. Intensity measurements on powders may be compared with those on single crystals by the application to the powder method data of the expression $A^2 = \frac{I}{j} \frac{\sin^2 \theta \cos \theta}{1 + \cos^2 2\theta}$ where I is the measured intensity of reflection and j is the number of planes belonging to the form which produces the reflection. Absolute measurements of the reflecting power have not been made.

Sodium Chloride.—Bragg, James and Bosanquet have measured the intensity of reflection of Rh Kα rays from single crystals of rock-salt, and have corrected their results for secondary extinction. Their method was unable to take account of primary extinction. The upper and lower curves in figure 1 represent their data, A^2 being plotted against $(h^2 + k^2 + l^2)$. The relative values of A^2 for the different planes are independent of wave-length, so the powder measurements made with Mo Kα radiation may be plotted in the same figure. Since primary extinction is proportional to the intensity and to the angle of reflection, any variation in extinction should be evidenced by a change in the relative values of A^2. Secondary extinction, which is proportional to the intensity of reflection only, should also be detected by a change in relative intensities, as is shown by the difference between the corrected and uncorrected curves of Bragg, James and Bosanquet.

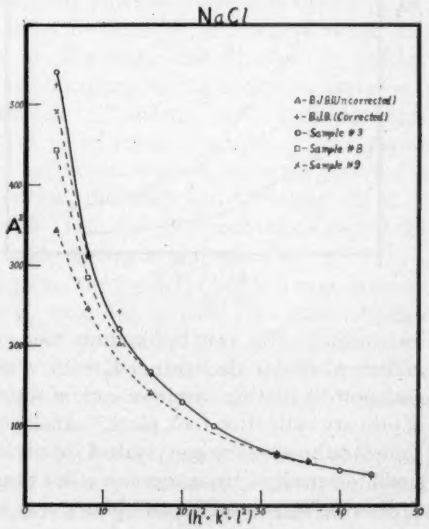


FIGURE 1

Sample 3 was a powder which had been ground until the particles were all smaller than 4×10^{-3} cm. The intensity measurements are seen to fall upon a smooth curve which coincides with that representing the data of

Bragg, James and Bosanquet, when the latter are corrected for secondary extinction. Some of the same powder was ground mechanically for several days, until the particle size was $1 \times 10^{-4} - 2 \times 10^{3}$ cm., and gave results identical with those of sample 3. Sample 6, a powder which had passed a 200-mesh sieve and been held back by a 325-mesh sieve, with particle size $4.4 - 7.4 \times 10^{-3}$ cm., was heated at 760° for 40 hours, then cooled gradually to room temperature. Because of the high vapor pressure at 760°, the crystals had grown. After having been reground to the original size, they gave results in agreement with sample 3. Samples 7 and 8 were from the same lot as 6 but were heated at a lower temperature, 700°, for 16 and 46 hours, respectively, then cooled slowly to room temperature. The part of the powder which remained within the original limits of size was investigated, care being taken to avoid mechanical strains. The results for these two samples were in close agreement, and are plotted in the figure. Evidently sufficient diffusion occurred within each powder particle to relieve some of the distortions, and a certain amount of primary extinction took place. Analogous results have been obtained by James[8] on heating single crystals of rock-salt. The powder particles were too small to permit of any noticeable effect from secondary extinction. Sample 8, after being ground for two hours, was reinvestigated, and then gave results in agreement with those from sample 3. Sample 9 was prepared by condensing the vapor of NaCl which had been vaporized from a hot platinum wire. Most of the particles were mere points when viewed through the oil immersion lens of a microscope, about 10^{-5} cm. in diameter, but a few were as large as 3×10^{-4} cm. The results show a slight amount of primary extinction, which modified the most intense reflection. Probably crystals formed from the vapor phase are more nearly perfect than those formed from solution, and even though most of the particles are so small that primary extinction is negligible, the few larger particles produce an

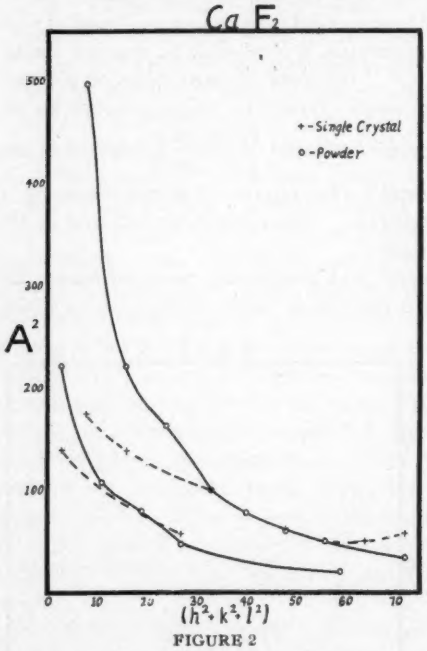


FIGURE 2

extinction the effect of which is measurable. Although sample 9 was later ground for two hours, there was no appreciable change in the intensities of reflection, possibly because the particles were so small as to undergo no strain during the grinding.

Calcium Fluoride.—MacInnes and Shedlovsky[9] have measured the relative intensities of reflection of Pd K rays from various planes of fluorite, but have made no attempt to correct their data for secondary extinction. Their data is plotted along with the powder method measurements in figure 2. The upper curve represents reflecting planes in which fluorine reenforces calcium, while the lower curve represents reflections from calcium planes alone. It is evident that extinction has here produced a notable effect. Three different samples of powdered CaF_2, all giving the same results, were used. Sample 1 was made by grinding a clear colorless fluorite crystal to $4.4-7.4 \times 10^{-3}$ cm. Sample 2 was from the same crystal but was further ground for 1 day to $4 \times 10^{-3}-1 \times 10^{-4}$ cm. Sample 3 was a pure precipitated powder with particles all smaller than 4×10^{-3} cm. It is possible that in addition to secondary extinction, which most certainly affected the results of MacInnes and Shedlovsky, primary extinction was also effective. As evidence to the contrary, however, it should be pointed out that both the (111) and the (400) reflections, which MacInnes and Shedlovsky found to be equal, are equal in the powder method results. Since primary extinction depends upon the angle of reflection as well as upon the intensity, the correction for reflections at the two different angles represented by (111) and (400) reflections should only be equal in case secondary extinction alone was operative.

Calcite.—Although it is difficult to make powder method measurements because of the great number of reflecting planes, measurements have been made with sufficient accuracy upon samples of ground calcite and precipitated calcium carbonate to show that the more intense reflections measured by W. H. Bragg[10] are several hundred per cent in error because of extinction.

Conclusions.—The conclusion to be drawn from this work, along with that of other investigators, is that primary extinction is absent in crystals that have been grown from solution and ground to less than 10^{-3} cm. in thickness and that measurements upon the intensity of reflection of X-rays by these powders may be used in the determination of electron distributions. Further, in crystals as poor as the alkali halides, single crystal measurements similar to those of Bragg, James and Bosanquet are satisfactory after correction for secondary extinction.

[1] NATIONAL RESEARCH FELLOW.

[2] W. L. Bragg, James and Bosanquet, *Phil. Mag.*, **44**, 433 (1922).
Havighurst, *Proc. Nat. Acad. Sci.*, **11**, 502, 507 (1925).
A. H. Compton, *Phys. Rev.*, **27**, 510 (1926).

[3] Darwin, *Phil. Mag.*, **43**, 800 (1922).

[4] Bragg, James and Bosanquet, *Ibid.*, **41**, 309; **42**, 1 (1921). What we call the (200) reflection is called by these authors, and by MacInnes and Shedlovsky, the (100) reflection.

[5] Mark, *Naturwissenschaften*, **13**, 1042 (1925).
Becker, *Physik. Z.*, **26**, 919 (1925).
Smekal, *Ibid.*, **26**, 707 (1925).

[6] Armstrong and Stifler, *J. Optical Soc. Am.*, **11**, 509 (1925).

[7] Brentano, *Proc. Phys. Soc., London*, **37**, 184 (1925).

[8] James, *Phil. Mag.*, **49**, 585 (1925).

[9] MacInnes and Shedlovsky, *Phys. Rev.*, **27**, 130 (1926).

[10] W. H. Bragg, *Phil. Trans. Roy. Soc.*, **A, 215**, 253 (1915).

THE INTENSITY OF REFLECTION OF X-RAYS BY LITHIUM, SODIUM AND CALCIUM FLUORIDES

By R. J. Havighurst[1]

Cruft Laboratory, Harvard University

Communicated May 5, 1926

The scattering power for X-rays of an atom in a crystal falls off with increasing angle of scattering in a manner that is determined by the electron distribution in the atom. In an effort to determine the electron distribution in certain light atoms, X-ray measurements have been made on the scattering power of powdered crystals of LiF, NaF and CaF_2.

Darwin's expression[2] for the power P diffracted in a cone of semi-apex angle 2θ by a bit of crystal powder of volume δV may be written

$$P = I \frac{j}{8} n^2 F^2 \lambda^3 \frac{e^4}{m^2 c^4} \frac{1 + \cos^2 2\theta}{\sin \theta} \delta V \qquad (1)$$

I is the power in the primary beam; n is the number of atoms per unit volume; j is the number of planes in the form under investigation; F^2 is a function of $\frac{\lambda}{\sin \theta}$ and of the arrangement of the electrons in the atoms, and includes the Debye temperature factor $e^{-\frac{b^2 \sin^2 \theta}{\lambda^2}}$. F, which is called the "atomic structure factor," approaches the number of electrons in the atom at small angles of reflection and theoretically should become equal to this number when $\sin \theta = 0$. All of the unknown quantities in the above expression except F may be measured, so that F may be calculated. Since the measurements reported in this note are only relative, we may dispense with the part of (1) which does not enter directly into the determination of the relative intensities. We then have

$$F^2 \propto \frac{P}{j} \frac{\sin\theta}{1+\cos^2 2\theta}. \qquad (2)$$

The power is measured by an ionization chamber at a distance b from the powder sample, with a slit of length l which is small compared with $b \sin 2\theta$. If the measurement is made so as to take in all of the angles in the neighborhood of 2θ at which any measurable power is diffracted, the power entering the ionization chamber is $P_s = \frac{Pl}{2\pi b \sin 2\theta}$.

Expression (2) then becomes

$$F^2 \propto \frac{P_s}{j} \frac{\sin^2\theta\cos\theta}{1+\cos^2 2\theta}. \qquad (3)$$

In order to get sufficient power in the scattered rays for accurate measurement at large angles, a focussing method described by Brentano[3] has been used. The primary beam from a Mo tube, filtered through enough ZrO_2 to reduce Mo K$(\gamma+\beta)$ to less than 0.4% of Mo K$(\alpha_1+\alpha_2)$, was reflected from the face of a briquet of the powder under investigation, the distance from focal spot to powder, a, and from powder to ionization chamber slit, b, determining the ratio $\frac{\sin\alpha}{\sin\beta} = \frac{a}{b}$, where α is the angle between the primary beam and the surface of the briquet, and $\beta = 2\theta - \alpha$. The tube was operated by a high potential storage battery[4] at 35,000 volts.

Relative intensities could be determined to an accuracy of considerably less than five per cent except in the case of the very weak reflections. The conceivable sources of systematic error have been investigated. It was found that the pressure used in making the briquet has a tendency to orient the surface layer of crystals if they are larger than 10^{-3} cm. in thickness. Consequently, the original surface of each sample was shaved off. The effect of particle size upon the intensity of reflection has been investigated (cf. preceding note) and it is believed that extinction was not operative in these experiments.

In figures 1 and 2, F curves for the different classes of crystal planes and their component atoms are plotted against $\sin\theta$. Since we are here dealing with relative values of the atomic structure factor, the ordinates will be called F'. However, an effort has been made to place these F' values on an approximate absolute scale. The author's F' curve for Na in NaCl was evaluated with the aid of the data of Bragg, James and Bosanquet on rock-salt,[5] and an F curve for Na was thus obtained and compared with the F' curve of Na from sodium fluoride. These two curves for Na could not be made to coincide, presumably because the electron distribution in the Na atom of NaCl is different from that of the Na atom in NaF. The

values of F for $\sin\theta = 0.20$ on the two curves were arbitrarily made equal. F' values for NaF were then determined and since all the crystals have fluorine in common, the F' curves for fluorine were made to agree as closely as possible. A two-fold assumption is therefore the basis of the valuation of these F' curves: (1) the scattering power of Na in NaCl and NaF is the same when $\sin\theta = 0.20$; (2) the scattering power of fluorine in LiF, NaF and CaF_2 is approximately the same at all angles. We may now consider the curves of figure 1.

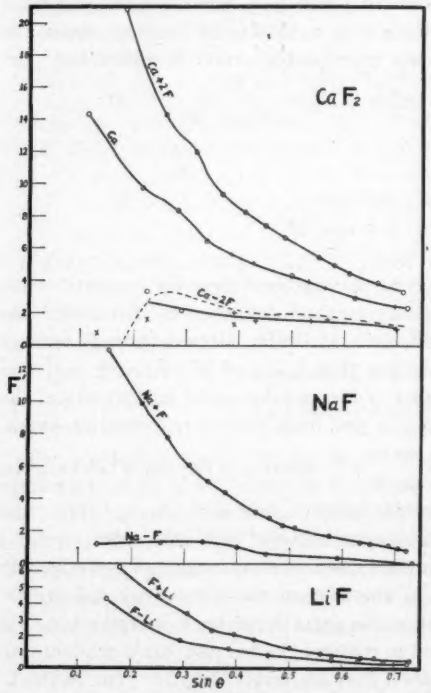


FIGURE 1

Lithium Fluoride.—The sample was carefully freed from Ca and Mg and several times reprecipitated. The upper curve represents planes with all even indices, in which Li and F reënforce each other; in the lower curve, for planes with all odd indices, Li interferes with F. Debye and Scherrer[6] have determined the relative intensities of reflection of the first nine planes of LiF by photometric study of powder method photographs. Their results fall reasonably well upon the curves of figure 1. They interpreted their work as proof that the crystal lattice consists of Li^+ and F^- ions, rather than of neutral atoms. We have seen that the F values should be equal to the number of electrons in the atoms composing the crystal planes at zero angle of reflection. Consequently the F + Li curve should approach the value 12, and the F − Li curve should approach the value 8 if the lattice consists of ions, or 6 if the lattice consists of neutral atoms. The corresponding ratios of $\frac{F + Li}{F - Li}$ are 1.5 and 2. Debye and Scherrer, by extrapolating the curve representing this ratio to zero angle of reflection, obtained a value of approximately 1.5. While there is abundant evidence from other sources that the lattice of the alkali halides is ionic, the shape of the F curves at small angles is entirely unknown to us and reliance can hardly be placed upon curves extrapolated into this region. The

ratio $\frac{F + Li}{F - Li}$ in figure 1 varies between 2 and 1.5, and seems to be approaching the smaller value at small angles.

Sodium Fluoride.—Samples from two different sources, each sample containing about 99% NaF, were investigated, with results which were in satisfactory agreement. The curve labelled Na−F is of particular interest because it represents the difference in scattering power of the Na^+ and F^- ions, on the assumption that the lattice is ionic. This curve should pass through zero at $\sin\theta = 0$. Since the ions have the same number of electrons, there seems at first to be no reason for assigning to Na^+ the larger scattering power. But Na^+, possessing a nuclear charge of +11, would bind its electrons more tightly than F^-, with a nuclear charge of +9. A system with electrons nearer its center has a larger scattering power through the angular domain covered by the curves of figure 1.

Calcium Fluoride.—Samples of ground fluorite, and a precipitated product containing less than half of one per cent impurity, gave similar results. Fluorite, with a crystal structure different from that of the alkali halides, gives reflections which fall into three classes, as shown in the figure. The Ca−2F reflections are very weak, and occur so close to the Ca reflections that their measurement is quite difficult. Only two direct measurements were made, shown by the circles on the curve; they are correct only in order of magnitude. The points marked + on the two lower curves were obtained by an indirect method. The combined intensity of a Ca reflection and its neighboring Ca−2F reflection was measured. Then an interpolated value from the Ca curve was used for the intensity of the Ca reflection and the amount left over from the total was assigned to the Ca−2F reflection. These values also are only approximate. The results of MacInnes and Shedlovsky[7] for the relative intensities of reflection from fluorite are affected by extinction (cf. preceding note), but the magnitudes of the weaker reflections have been located approximately in the figure, and marked X. The broken curve is that obtained for Ca−2F from the two upper curves. If the Ca^{++} ion, with 18 electrons, is assumed to be present in the lattice along with $2F^-$ ions, with 10 electrons apiece, the curve representing Ca−2F must drop to zero and then rise as a 2F−Ca curve to a value of 2 when $\sin\theta = 0$. Here we see that Ca^{++}, with an excess positive charge, has drawn its electrons in and consequently scatters X-rays more strongly than the two F^- ions, which possess an excess of negative charge.

Atomic Scattering Curves.—In figure 2 are the atomic scattering curves obtained from the curves of figure 1. Of considerable interest is the fact that the curves for fluorine from different compounds are nearly alike. Provided that these curves do not differ on an absolute scale, the conclusion is that the fluorine ion is in a force field of the same magnitude in all

three crystals. As has already been pointed out, the Na curves from NaF and NaCl do not coincide, probably because the chemical forces are of different magnitudes in the two crystals. Examination of the heats of formation of the compounds in question gives qualitative support to this explanation. The fluorine curve of CaF_2 possesses a slight hump, as do the Ca and Li curves. If experimental error is ruled out, the humps indicate either a definite shell-like structure of the electrons, or a lack of spherical symmetry in the atoms. Li^+, with only two K electrons, could hardly give anything except a smooth scattering curve unless the system lacks

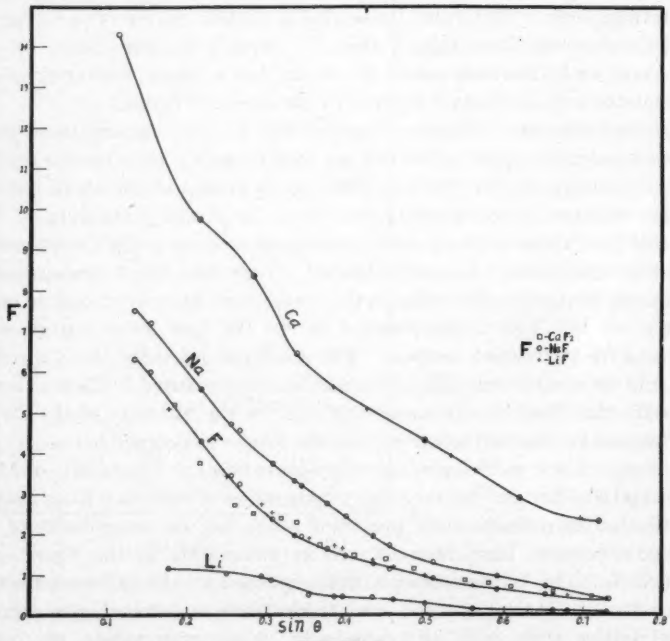


FIGURE 2

spherical symmetry. The electron distribution curves which are now being worked out for these crystals by a method involving a Fourier analysis[8] should help to answer many of the questions raised by the appearance of these curves.

The atomic scattering curves show clearly that the assignment of scattering power to atoms in accordance with their total number of electrons is only a crude approximation. A similar variation of scattering power with angle of scattering is predicated for all ions by such a procedure. Especially uncertain is this assignment of scattering power in the case

of oppositely charged ions (the usual case in crystal analysis) because the excess of positive or negative charge produces a tightening or loosening of the electron atmosphere of the original atom which profoundly affects its scattering properties.

[1] NATIONAL RESEARCH FELLOW.
[2] Darwin, *Phil. Mag.*, **43**, 800 (1922).
[3] Brentano, *Proc. Phys. Soc., London*, **37**, 184 (1925).
[4] Armstrong and Stifler, *J. Optical Soc. Am.*, **11**, 509 (1925).
[5] W. L. Bragg, James and Bosanquet, *Phil. Mag.*, **41**, 309; **42**, 1 (1921).
[6] Debye and Scherrer, *Physik. Z.*, **19**, 474 (1918).
[7] MacInnes and Shedlovsky, *Phys. Rev.*, **27**, 130 (1926).
[8] Havighurst, *Proc. Nat. Acad. Sci.*, **11**, 502, 507 (1925).

NOTE ON THE POSTULATES OF THE MATRIX QUANTUM DYNAMICS

BY J. H. VAN VLECK

DEPARTMENT OF PHYSICS, UNIVERSITY OF MINNESOTA

Communicated April 22, 1926

Through the researches of Born, Heisenberg and Jordan,[1,2,3] and of Dirac,[4,5] the dynamics of the quantum theory have been formulated in terms of matrices. The fundamental postulates made by Born, Heisenberg and Jordan are the Ritz combination principle

$$\nu(nm) + \nu(mk) = \nu(nk) \tag{1}$$

the Hamiltonian canonical equations[6]

$$\dot{\mathbf{q}}_k = \frac{\partial \mathbf{H}}{\partial \mathbf{p}_k}, \qquad \dot{\mathbf{p}}_k = -\frac{\partial \mathbf{H}}{\partial \mathbf{q}_k}, \tag{2}$$

the quantum conditions,

$$\mathbf{p}_k\mathbf{q}_k - \mathbf{q}_k\mathbf{p}_k = \frac{h}{2\pi i}\mathbf{1} \tag{3}$$

and the commutability relations

$$\begin{aligned} \mathbf{p}_l\mathbf{q}_k - \mathbf{q}_k\mathbf{p}_l &= 0, \qquad (l \neq k), \\ \mathbf{q}_l\mathbf{q}_k - \mathbf{q}_k\mathbf{q}_l &= 0, \qquad \mathbf{p}_l\mathbf{p}_k - \mathbf{p}_k\mathbf{p}_l = 0. \end{aligned} \tag{4}$$

Here $\mathbf{H}$ is the energy (Hamiltonian function) and the $\mathbf{q}$'s and $\mathbf{p}$'s are coördinates and momenta. We shall suppose that there are s degrees of freedom, so that the subscripts k and l range from 1 to s. All expressions

printed in bold-face type are matrices, as we follow Born's notation throughout. From (1), (2), (3) and (4) Born, Dirac and others deduce the Bohr frequency condition

$$h\nu(nm) = H(nn) - H(mm) \tag{5}$$

and the conservation of energy

$$\dot{\mathbf{H}} = 0. \tag{6}$$

The purpose of the present note is to call attention to the fact that (2), (4), (5) and (6) can be taken as the basic postulates of the new theory instead of (1), (2), (3) and (4). The fundamental postulates are then the canonical equations, the commutability of the order of multiplication in all the products $\mathbf{q}_k\mathbf{q}_l$, $\mathbf{p}_k\mathbf{p}_l$, $\mathbf{q}_k\mathbf{p}_l$, except $\mathbf{q}_k\mathbf{p}_k$, the Bohr frequency condition, and the conservation of energy. In other words, if we accept (2) and (4), the equations (1) and (3) are not merely *sufficient* (as proved by Born, Heisenberg and Jordan[3]) but also *necessary* for the validity of the Bohr frequency condition and the conservation of energy. This is not surprising, but is perhaps worth noting inasmuch as it does not appear to have been specifically mentioned in the various papers previously written on the matrix dynamics. In fact, the Bohr frequency condition and the conservation of energy are more familiar concepts than the quantum conditions (3) and so may appeal to some readers as more lucid postulates than (1) and (3). One must still make the postulate (4) that the order of multiplication is commutative in all products of the form $\mathbf{q}_k\mathbf{q}_l$, $\mathbf{p}_k\mathbf{p}_l$, $\mathbf{q}_k\mathbf{p}_l$, except $\mathbf{q}_k\mathbf{p}_k$, and unfortunately the basis for this assumption is not apparent unless, following Dirac,[4] we establish a correspondence with the Poisson's bracket expressions of ordinary dynamical contact transformation theory. In systems with one degree of freedom, however, there is only one $\mathbf{q}$ and $\mathbf{p}$, and one can then dispense with (4) entirely, so that the fundamental postulates take the very simple form (2), (5) and (6).

Proof.—The Ritz combination principle (1) is an obvious consequence of the Bohr frequency condition (5). To show that postulates (1) and (3) can be replaced by (5) and (6) we must in addition deduce (3) from (2), (4), (5) and (6). Our proof will be very similar to Born, Heisenberg and Jordan's converse derivation[3] of (5) and (6) from (1), (2), (3) and (4). By virtue of the Bohr frequency condition and the constancy of energy, we have

$$h\dot{\mathbf{q}}_k = 2\pi i(\mathbf{H}\mathbf{q}_k - \mathbf{q}_k\mathbf{H}). \tag{7}$$

This is a well-known result. It follows directly inasmuch as an element of the matrix on the left-hand side of (7) is

$$h.2\pi i\nu(nm)a(nm)\exp.2\pi i\nu(nm)t$$

while an element of the right-hand side is

$$2\pi i(H(nn) - H(mm))a(nm)\exp.2\pi i\nu(nm)t$$

in consequence of the rules for multiplication of matrices and the fact that $\mathbf{H}$ is a diagonal matrix because of (6). Here $a(nm)\exp.2\pi i\nu(nm)t$ denotes an element of the matrix $\mathbf{q}_k$. Equation (7) is thus equivalent to (5) whenever the $\mathbf{q}_k$ and $\mathbf{p}_k$ can be represented by matrices, as in multiply periodic systems. Equation (7) is, however, more general than (5) inasmuch as (7) is applied by Born and Wiener[7] and by Dirac[5] even to "aperiodic" motions which cannot be represented by ordinary matrices. The derivation which we shall give of (3) applies to aperiodic systems if we take (7) as a fundamental postulate in place of (5) and (6).

Comparing (2) and (7) we see that

$$\mathbf{H}\mathbf{q}_k - \mathbf{q}_k\mathbf{H} = \frac{\partial \mathbf{H}}{\partial \mathbf{p}_k}\frac{h}{2\pi i}. \tag{8}$$

We shall next establish the equation

$$\mathbf{f}\mathbf{q}_k - \mathbf{q}_k\mathbf{f} = \frac{\partial \mathbf{f}}{\partial \mathbf{p}_k}\frac{h}{2\pi i} \qquad (k = 1,..,s) \tag{9}$$

where $\mathbf{f}$ is any function of the $\mathbf{p}$'s and $\mathbf{q}$'s. This equation was deduced by Born, Heisenberg and Jordan from (1), (2), (3) and (4), whereas we must deduce it from (2), (4) and (7). Before proceeding to the proof we must define what is meant by the derivative of one matrix with respect to another, as such a derivative is involved on the right-hand side of (9). We shall assume that the operation of matrix differentiation is defined in such a way that

$$\begin{gathered}\frac{\partial \mathbf{p}_l}{\partial \mathbf{p}_k} = 0\ (l \neq k), \quad \frac{\partial \mathbf{p}_k}{\partial \mathbf{p}_k} = \mathbf{1}, \quad \frac{\partial \mathbf{q}_l}{\partial \mathbf{p}_k} = 0, \\ \frac{\partial}{\partial \mathbf{p}_k}(\mathbf{a} + \mathbf{b}) = \frac{\partial \mathbf{a}}{\partial \mathbf{p}_k} + \frac{\partial \mathbf{b}}{\partial \mathbf{p}_k}, \quad \frac{\partial}{\partial \mathbf{p}_k}(\mathbf{ab}) = \frac{\partial \mathbf{a}}{\partial \mathbf{p}_k}\mathbf{b} + \mathbf{a}\frac{\partial \mathbf{b}}{\partial \mathbf{p}_k},\end{gathered} \tag{10}$$

where $\mathbf{a}$ and $\mathbf{b}$ are any two functions of the coördinates and momenta. This procedure is tantamount to taking the derivative to be a "differential quotient of the first kind" in the terminology of Born, Heisenberg and Jordan.[3] In fact the "differential quotients of the first kind" are perhaps more readily understood if defined by (10) rather than by the alternative limiting process used by Born, Heisenberg and Jordan.[3]

Now by (4) and (8), equation (9) is obviously valid in the particular case that $\mathbf{f}$ is any one of the $2s$ expressions

$$\mathbf{p}_1, \ldots, \mathbf{p}_{k-1}, \mathbf{p}_{k+1}, \ldots, \mathbf{p}_s, \mathbf{q}_1, \ldots, \mathbf{q}_s, \mathbf{H}. \tag{11}$$

Consequently (9) is valid when $\mathbf{f}$ is any function of the $2s$ arguments (11). To establish this we need only note, with Born, Heisenberg and Jordan,[3] that functions of matrices are defined by successive applications of the operations of addition and multiplication, and that if (9) holds for $\mathbf{f} = \mathbf{a}$ and $\mathbf{f} = \mathbf{b}$, it also holds for $\mathbf{f} = \mathbf{a} + \mathbf{b}$ and $\mathbf{f} = \mathbf{ab}$. In the case of $\mathbf{f} = \mathbf{a} + \mathbf{b}$ the truth of this statement is obvious, while for $\mathbf{f} = \mathbf{ab}$ we have (cf. Born[3]):

$$\mathbf{abq}_k - \mathbf{q}_k\mathbf{ab} = \mathbf{a}(\mathbf{bq}_k - \mathbf{q}_k\mathbf{b}) + (\mathbf{aq}_k - \mathbf{q}_k\mathbf{a})\mathbf{b}$$
$$= \left(\mathbf{a}\frac{\partial \mathbf{b}}{\partial \mathbf{p}_k} + \frac{\partial \mathbf{a}}{\partial \mathbf{p}_k}\mathbf{b}\right)\frac{h}{2\pi i} = \frac{\partial}{\partial \mathbf{p}_k}(\mathbf{ab})\frac{h}{2\pi i} \text{ by (10).}$$

Since (9) thus holds for any function of the $2s$ arguments (11) it also holds for any function of the $2s$ coördinates $\mathbf{q}_l$ and $\mathbf{p}_l$ ($l = 1, \ldots s$). This follows since by solving the energy equation

$$\mathbf{H} = \mathbf{H}(\mathbf{q}_1, \ldots, \mathbf{q}_s, \mathbf{p}_1, \ldots \mathbf{p}_s)$$

for the variable $\mathbf{p}_k$, we can express $\mathbf{p}_k$ in terms of the arguments (11). Thus $\mathbf{p}_k$ can be eliminated and the arguments of $\mathbf{f}$ always regarded as (11).

As we have shown (9) holds for any function $\mathbf{f}$, we can, in particular, take $\mathbf{f} = \mathbf{p}_k$. This gives immediately the desired result (3), Q.E.D. We could also easily establish the relations

$$\mathbf{p}_k\mathbf{f} - \mathbf{f}\mathbf{p}_k = \frac{\partial \mathbf{f}}{\partial \mathbf{q}_k}\frac{h}{2\pi i}$$

which are analogous to (9) but which are not needed in proving (3).

The writer wishes to express his thanks to Prof. Tate for suggesting the question discussed in the present note.

[1] W. Heisenberg, *Zeit. f. Phys.*, **33**, 879 (1925).

[2] M. Born and P. Jordan, *Ibid.*, **34**, 858 (1925).

[3] M. Born, W. Heisenberg and P. Jordan, *Ibid.*, **35**, 557 (1926).

[4] P. A. M. Dirac, *Proc. Roy. Soc., London*, **A, 109**, 642 (1925).

[5] P. A. M. Dirac, *Ibid.*, **A, 110**, 561 (1925).

[6] Born and Jordan[2] deduce the canonical equations from a variation principle, but the canonical equations themselves can, of course, be taken equally well as fundamental postulates, as in the present note.

[7] M. Born and N. Wiener, *Jour. of Math. and Phys., Mass. Inst. of Tech.*, **5**, 84 (1926), or *Zeit. f. Phys.*, **36**, 174 (1926).

CONFORMAL INVARIANTS

BY JOSEPH MILLER THOMAS*

PRINCETON UNIVERSITY

Communicated April 1, 1926

1. I have previously[1] pointed out the existence of a conformal invariant, the conformal connection, which in the problem of the conformal correspondence of Riemann spaces plays a rôle similar to that of the Christoffel symbols in the problem of the equivalence of two quadratic differential forms, and by expressing integrability conditions of its law of transformation given a derivation of the conformal curvature tensor discovered by Weyl. By continuing the process of elimination employed in the former paper, I now show how a complete set of integrability conditions for the equations of transformation of the conformal connection can be obtained. These integrability conditions express the laws of transformation of a set of conformal invariants. In addition to the first partial derivatives of the old variables with respect to the new, they involve certain combinations of second derivatives, but no derivatives of higher order than the second. Consequently the resulting invariants are not in general tensors. They form, nevertheless, a basis for stating an equivalence theorem. The sequence of invariants is composed of two invariants and their successive conformal derivatives, the conformal derivative being defined as the ordinary derivative evaluated in a certain conformal coördinate system.

2. From the results of the former paper the conformal connection is given by

$$K^i_{jk} = \left\{ {i \atop jk} \right\} - (\delta^i_j \left\{ {l \atop lk} \right\} + \delta^i_k \left\{ {l \atop lj} \right\} - g^{ip} g_{jk} \left\{ {l \atop lp} \right\})/n, \qquad (2.1)$$

and its law of transformation is

$$u^i_{jk} = \overline{K}^l_{jk} u^i_l - K^i_{lp} u^l_j u^p_k + u^i_j \psi_k + u^i_k \psi_j - \bar{g}_{jk} \bar{g}^{ip} u^i_l \psi_p. \qquad (2.2)$$

In these formulas $u^i_j = \partial x^i / \partial \bar{x}^j$, $u^i_{jk} = \partial^2 x^i / \partial \bar{x}^j \partial \bar{x}^k$, $\Delta(x, \bar{x}) = \left| \partial x / \partial \bar{x} \right|$, $\psi = \log \Delta / n$, $\psi_i = \partial \psi / \partial \bar{x}^i$.

If we differentiate equations (2.2) with respect to $\bar{x}^l$, eliminate u^i_{jk} by means of (2.2), interchange k and l, and subtract, we find as in the former paper equations equivalent to

$$\overline{F}^i_{jkl} = F^p_{qsl} U^i_p u^q_j u^s_k u^t_l - \delta^i_k c_{jl} + \delta^i_l c_{jk} - \bar{g}_{jl} c^i_k + \bar{g}_{jk} c^i_l + (\delta^i_l \bar{g}_{jk} - \delta^i_k \bar{g}_{jl}) \psi_p \psi^p, \qquad (2.3)$$

where

$$F^i_{jkl} = \frac{\partial K^i_{jk}}{\partial x^l} - \frac{\partial K^i_{jl}}{\partial x^k} + K^p_{jk} K^i_{pl} - K^p_{jl} K^i_{pk},$$

$$c_{jk} = \psi_{jk} - \psi_j \psi_k - \psi_p \overline{K}^p_{jk},$$

and $c^i_k = \bar{g}^{ip}\, c_{pk}$, $\psi^p = \bar{g}^{pq}\, \psi_q$, $\psi_{jk} = \partial\psi_j/\partial\bar{x}^k$, $U^i_j = \partial\bar{x}^i/\partial x^j$. Contraction for i and l in (2.3) gives

$$\bar{F}_{jk} = F_{pq}\, u^p_j\, u^q_k + (n-2)c_{jk} + \bar{g}_{jk}\, c + (n-1)\bar{g}_{jk}\, \psi_p\, \psi^p, \quad (2.4)$$

where $F_{pq} = F^s_{pqs}$, $c = \bar{g}^{pq}\, c_{pq}$. Multiplication of (2.4) by $\bar{g}^{jk}$ gives

$$\bar{F} = F + 2(n-1)c + n(n-1)\psi_p\, \psi^p, \quad (2.5)$$

where $F = g^{pq}\, F_{pq}$. From (2.4) and (2.5) we find

$$\psi_{jk} = \bar{f}_{jk} - f_{pq}\, u^p_j\, u^q_k - \tfrac{1}{2}\bar{g}_{jk}\, \bar{g}^{pq}\, \psi_p\, \psi_q + \bar{K}^p_{jk}\, \psi_p + \psi_j\, \psi_k, \quad (2.6)$$

$$f_{jk} = [F_{jk} - g_{jk}\, F/2(n-1)]/(n-2).$$

Elimination of ψ_{jk} from equations (2.3) by means of (2.6) leads to the equations,

$$\bar{C}^p_{jkl}\, u^i_p = C^i_{pqs}\, u^p_j\, u^q_k\, u^s_l, \quad (2.7)$$

$$C^i_{jkl} = F^i_{jkl} + \delta^i_k f_{jl} - \delta^i_l f_{jk} + g_{jl} f^i_k - g_{jk} f^i_l, \quad (2.8)$$

in which $f^i_k = g^{ip}\, f_{pk}$. Hence the first invariant of the set happens to be a tensor. It is, except for the factor $(2-n)$, the conformal curvature tensor of Weyl.[2]

The first invariant arising from the integrability conditions of (2.6) is

$$f_{jkl} = \partial f_{jk}/\partial x^l - \partial f_{jl}/\partial x^k + K^p_{jk} f_{pl} - K^p_{jl} f_{pk},$$

and its law of transformation is

$$f_{jkl} = f_{pqs}\, u^p_j\, u^q_k\, u^s_l - \psi_p\, C^p_{jkl}. \quad (2.9)$$

We shall call f_{jkl} the *conformal covariant*. From equations (2.9) follows:

THEOREM 1. *The conformal covariant is a tensor for spaces whose conformal curvature tensor vanishes, and only for such spaces.*

If we differentiate (2.7) and (2.9) successively and eliminate each time the second derivatives which occur explicitly by means of (2.2) and (2.6), we get the desired set of conformal invariants together with their laws of transformation. The latter involve no derivatives of the coördinates x except $\partial x/\partial\bar{x}$ and $\partial\psi/\partial\bar{x}$.

The first invariant arising from the differentiation of (2.7) is

$$C^i_{jkl/p} = \partial C^i_{jkl}/\partial x^p + C^q_{jkl}\, K^i_{qp} - C^i_{qkl}\, K^q_{jp} - C^i_{jql}\, K^q_{kp} - C^i_{jkq}\, K^q_{lp}, \quad (2.10)$$

that is, it is an invariant formed from C^i_{jkl} in the same way as its ordinary covariant derivative, but with the K's replacing the Christoffel symbols. This invariant is not a tensor since its law of transformation involves the ψ_i.

Let us denote the formal expression for the covariant derivative of any quantity with respect to the K's by a notation like that in (2.10). Since

equations (2.5) of the former paper show that $(g_{jl}\, g^{iq})_{/p} = 0$, we find from (2.8)

$$C^i_{jkl/p} = F^i_{jkl/p} + \delta^i_k f_{jl/p} - \delta^i_l f_{jk/p} + g_{jl}\, g^{iq} f_{qk/p} - g_{jk}\, g^{iq} f_{ql/p}.$$

Also by the use of the relations

$$F^i_{jkl/p} + F^i_{jlp/k} + F^i_{jpk/l} = 0,$$

$$f_{ijk} = f_{ij/k} - f_{ik/j},$$

we get

$$C^i_{jkl/p} + C^i_{jlp/k} + C^i_{jpk/l} = \delta^i_k f_{jl/p} + \delta^i_l f_{jp/k} + \delta^i_p f_{jkl} + g^{iq}(g_{jl} f_{qk/p} + g_{jp} f_{qlk} + g_{jk} f_{qpl}),$$

whence by contraction

$$C^p_{jkl/p} = (n-3) f_{jkl}. \tag{2.11}$$

3. A system of conformal coördinates y associated with a given coordinate system x and a point x_0 is defined by the equations

$$x^i = x^i_0 + y^i - \tfrac{1}{2!}(K^i_{jk})_0\, y^j y^k - \tfrac{1}{3!}(K^i_{jkl})_0\, y^j y^k y^l, \tag{3.1}$$

where

$$K^i_{jkl} = \tfrac{1}{3} P\left(\frac{\partial K^i_{jk}}{\partial x^l} - 2K^p_{jk} K^i_{pl} + 2\delta^i_l f_{jk} - g_{jk} f^i_l\right),$$

the symbol P denoting the sum of the terms obtained from those in the parentheses by cyclic permutation of the subscripts.

We define the *conformal derivative* of an invariant whose transformation law involves only $\partial x/\partial \bar{x}$ and $\partial \psi/\partial \bar{x}$ as the ordinary derivative evaluated in the system of coördinates (3.1) at the origin. This derivative will have only $\partial x/\partial \bar{x}$ and $\partial \psi/\partial \bar{x}$ in its law of transformation, that is, it is of the same class as the invariant which was differentiated.

The set of conformal invariants obtained in §2 are the successive conformal derivatives of C^i_{jkl} and f_{jkl}. To prove this for a given one of the invariants, we differentiate the law of transformation of the invariant immediately preceding it in the sequence, then regard $\bar{x}$ as the coördinates of (3.1) and evaluate at the origin where the following conditions hold:

$$\frac{\partial x^i}{\partial y^j} = \delta^i_j, \qquad \frac{\partial^2 x^i}{\partial y^j \partial y^k} = -K^i_{jk}, \qquad \psi_i = 0, \qquad \overline{K}^i_{jk} = 0, \qquad \overline{f}_{ij} = 0.$$

By a complete set of invariants we mean one in terms of which an equivalence theorem can be stated. In order to make the above set of invariants complete, we add to it the conformal relative tensor[3] $g_{ij}/g^{\frac{1}{n}}$ where $g = |g_{ij}|$. The statement and proof of the conformal equivalence theorem are omitted here because they are almost identical with those of the cor-

responding projective theorem. The latter is stated and proved in an article on "Projective Invariants of Affine Geometry of Paths," by Professor Veblen and the writer in volume 27 (1926) of the *Annals of Mathematics*.

If we admit that such a theorem can be proved, we have the following.

THEOREM 2. *A complete set of conformal invariants for a Riemann space consists of the relative tensor* $g_{ij}/g^{\frac{1}{n}}$, *the conformal curvature tensor and the conformal covariant together with their successive conformal derivatives.*

If $n > 3$, it follows from (2.11) that the conformal covariant can be omitted from the list. For $n = 3$, the conformal curvature tensor vanishes identically and can be omitted.

4. By a conformally flat space is meant one which can be represented conformally on a euclidean space. A necessary condition for such a space is the existence of a preferred coördinate system in which the conformal connection vanishes identically, this coördinate system being cartesian for the euclidean space. The condition is also sufficient. From (2.1) we get on the assumption that $K^i_{jk} = 0$,

$$n\left\{\begin{matrix} i \\ jk \end{matrix}\right\} = \delta^i_j\left\{\begin{matrix} l \\ lk \end{matrix}\right\} + \delta^i_k\left\{\begin{matrix} l \\ lj \end{matrix}\right\} - g^{ip}\, g_{jk}\left\{\begin{matrix} l \\ lp \end{matrix}\right\},$$

or

$$n[jk,i] = g_{ij}\left\{\begin{matrix} l \\ lk \end{matrix}\right\} + g_{ik}\left\{\begin{matrix} l \\ lj \end{matrix}\right\} - g_{jk}\left\{\begin{matrix} l \\ li \end{matrix}\right\}.$$

We readily find from the last equations

$$\frac{\partial \log g_{ij}}{\partial x^k} = \frac{1}{n}\frac{\partial \log g}{\partial x^k},$$

whence $g_{ij} = g^{\frac{1}{n}} a_{ij}$, the a's being constants. We therefore have $g_{ij}\, dx^i\, dx^j = g^{\frac{1}{n}} a_{ij}\, dx^i\, dx^j$, equations which show that the space can be represented conformally on a euclidean space.

Since any surface is conformally flat, we assume from now on that $n \geq 3$.

Suppose that there exists a preferred coördinate system $\bar{x}$ such that $\overline{K}^i_{jk} = 0$. The invariants $\bar{f}$ and $\overline{C}$ vanish in such a coördinate system and hence, see (2.7) and (2.9), in any coördinate system. Conversely, if f and C vanish in some coördinate system x, then equations (2.2) and (2.6) in which $\overline{K}^i_{jk}$ have been put equal to zero are completely integrable and define a preferred coördinate system $\bar{x}$ in which $\overline{K}$ vanishes identically.

From (2.11) it is seen that the vanishing of C for $n > 3$ implies that of f, whereas for $n = 3$, C is identically zero and the condition for flatness is the vanishing of f. This was first noticed by J. A. Schouten in somewhat different form.[4] Schouten's conditions are

$$C_{jkl} = 0, \qquad (4.1)$$

$$L_{jk,l} - L_{jl,k} = 0,$$

where in general

$$L_{jk,l} - L_{jl,k} = (n-2)f_{jkl} + (n-2)\left\{ {a \atop qp} \right\} C^{p}_{jkl}/n. \qquad (4.2)$$

My conditions are therefore equivalent to Schouten's. It is to be noticed, however, that the former are expressed in terms of invariants which are obviously conformal in character. Schouten's invariant (4.2) is conformal only when taken in conjunction with (4.1) and even then its conformal character is not obvious.

We now state in terms of the invariants of the present paper two theorems about conformally flat spaces. The first embodies a theorem discovered by Cotton,[5] the second was discovered by Schouten.

THEOREM 3. *For a Riemann space of three dimensions the conformal covariant is a tensor and its vanishing is a necessary and sufficient condition that the space be conformally flat.*

THEOREM 4. *A necessary and sufficient condition that a Riemann space of more than three dimensions be conformally flat is the vanishing of the conformal curvature tensor.*

* NATIONAL RESEARCH FELLOW IN MATHEMATICS.

[1] These PROCEEDINGS, **11** (1925), 257–9.

[2] H. Weyl, *Math. Zeit.*, **2** (1918), 404.

[3] That this relative tensor plays a leading rôle in the conformal geometry was first pointed out by T. Y. Thomas, these PROCEEDINGS, **11** (1925), 722–5.

[4] *Math. Zeit.*, **11** (1921), 83; *der Ricci-Kalkül*, Berlin, 1924, p. 170.

[5] E. Cotton, *Ann. Fac. Sci. Toulouse*, **1** (1899), 412.

GENERALIZATION OF KRONECKER'S RELATION AMONG THE MINORS OF A SYMMETRIC DETERMINANT

BY HELEN BARTON

JOHNS HOPKINS UNIVERSITY

Communicated March 30, 1926

Since 1882, when Kronecker stated, without proof, that there exists a certain linear relation among the minors of a symmetric determinant, there have been many and varied proofs. It is the purpose of this paper to show that this relation is but a special case of a more general relation.[1]

The theorem as given by Kronecker is: That among the minors of order m of a symmetric determinant of order $2m$ there exists the following linear relation

$$|a_{gh}| = \sum_r |a_{ik}|$$

where $g = 1, 2, \ldots\ldots m \qquad h = m+1, m+2, \ldots\ldots 2m$

$i = 1, 2, \ldots m-1, r \qquad M = m+1, m+2, \ldots r-1, m, r+1, \ldots 2m$

$r = m+1, m+2, \ldots\ldots\ldots 2m.$

CASE I. To prove this relation, let us consider the following expression, which involves the Generalized Kronecker Symbol.[2]

$$\begin{bmatrix} \beta_m & \beta_{m+1} & \cdots & \beta_{2m} \\ m & m+1 & \cdots & 2m \end{bmatrix} a_{1\beta_{m+1}} a_{2\beta_{m+2}} \cdots\cdots a_{m-1\,\beta_{2m-1}} a_{\beta_m\,\beta_{2m}}.$$

For any given set of values for $\beta_{m+1} \ldots \beta_{2m-1}$; (for example, b_m, b_{m+1}, $\ldots b_{m\backslash+r} \ldots b_{m\backslash+s} \ldots b_{2m}$—which means a complete set of b's exclusive of the $2b$'s which are crossed out) β_m and β_{2m} must each take either of the values b_{m+r} or b_{m+s}. Therefore we get two expressions, namely,

$$\begin{bmatrix} b_{m+r} & b_{m+1} & \cdots & b_{2m} & b_{m+s} \\ m & m+1 & \cdots\cdots & 2m & \end{bmatrix} a_{1b_{m+1}} \cdots\cdots a_{m-1\,b_{2m-1}} a_{b_{m+r}\,b_{m+s}}$$

and $$\begin{bmatrix} b_{m+s} & b_{m+1} & \cdots & b_{m+r} \\ m & m+1 & \cdots & 2m \end{bmatrix} a_{1b_{m+1}} \cdots\cdots a_{m-1\,b_{2m-1}} a_{b_{m+s}\,b_{m+r}}$$

which are equal in value, but opposite in sign; hence their sum equals zero. Since this is true for all combinations $b_{m+1}, \ldots b_{m\backslash+r} \ldots b_{m\backslash+s} \ldots b_{2m}$ the total expression will consist in a pairing off of terms, each pair reducing to zero, and thereby causing the whole expression to equal zero.

But the expression may be written as a summation in a slightly different form. Consider β_m as it assumes in succession the $m+1$ values $m, m+1, \ldots 2m$; then corresponding to each value of β_m, $\beta_{m+1} \ldots \beta_{2m}$ will assume all possible arrangements of the remaining values. But each of these expressions is a determinant of the mth order. Therefore, if we take into account also the sign factor in each case, as given by the Generalized Kronecker Symbol, the truth of the following is evident.

$$\begin{bmatrix} \beta_m & \beta_{m+1} & \cdots & \beta_{2m} \\ m & m+1 & \cdots & 2m \end{bmatrix} a_{1\beta_{m+1}} a_{2\beta_{m+2}} \cdots\cdots a_{\beta_m\,\beta_{2m}} =$$

$$\left| a_{1,\,m+1}\, a_{2,\,m+2} \cdots\cdots a_{m,\,2m} \right| - \left| a_{1,\,m}\, a_{2,\,m+2} \cdots a_{m-1,\,2m-1}\, a_{m+1,\,2m} \right| -$$

$$\left| a_{1,\,m+1}\, a_{2,\,m} \cdots\cdots a_{m+2,\,2m} \right| \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$- \left| a_{1,\,m+1}\, a_{2,\,m+2} \cdots\cdots\cdots\cdots a_{m-1,\,2m-1}\, a_{2m,\,m} \right| = 0,$$

i.e., $\left| a_{gh} \right| = \sum_r \left| a_{ik} \right|$.

We now proceed to a slightly more general case, where instead of one summation symbol, occurring in the first set of values (i.e., from 1 to m) we have two.

CASE II. We now consider the expression

$$\begin{bmatrix} \beta_{m-1} & \beta_m & \beta_{m+1} & \cdots & \beta_{2m} \\ m-1 & m & m+1 & \cdots & 2m \end{bmatrix} a_{1\beta_{m+1}} a_{2\beta_{m+2}} \cdots a_{m-2\,\beta_{2m-2}} a_{\beta_{m-1}\beta_{2m-1}} a_{\beta_m \beta_{2m}}.$$

This expression reduces to zero, since as before $a_{\beta_m \beta_{2m}}$ causes a pairing off of terms, so that the terms of each pair cancel one another.

In the preceding case we arranged the expression as a summation of determinants, obtained by allowing β_m to assume in succession all values from m to $2m$; so here we will group according to the values assumed by β_{m-1} and β_m. For each set of values assumed by β_{m-1} and β_m the remaining β's will take all possible arrangements of the remaining values from $(m-1, m \ldots 2m)$ and as before we shall have a summation of various determinants. But for any one set of values assumed by β_{m-1} and β_m; $-b_{m+r}$ and b_{m+s}, for example; there are two possible arrangements, depending upon whether B_{m-1} assumes the value b_{m+r} or b_{m+s}. Since a rearrangement of values assigned to β_{m-1} and β_m means not only a change of sign in the sign factor but also a change of sign of each term, it follows that the total result of such an interchange leaves the expression unchanged. Hence, there will be a complete duplication of determinants, and we may remove the factor $2!$, provided we note that in our summation we consider only combinations of values for β_{m-1} and β_m and not permutations.

In grouping these various determinants which result, we note that they fall into the following groups:

1. β_{m-1} and β_m assume the two values $m-1$ and m in the first set. But this is merely one combination of values, hence we represent the determinant which results by $|a_{gh}|$.

2. β_{m-1} assumes values in the first set and β_m in the second and vice versa. In this case, suppose β_{m-1} assumes the value $m-1$; then let β_m assume all values from $m+1$ to $2m$. This is nothing but case I and the sum of the determinants so obtained $= |a_{gh}|$.

In an exactly similar way if β_m remains fixed, with value m and we let β_{m-1} assume all the remaining values, we again get a sum of determinants which equals $|a_{gh}|$.

3. Both β_{m-1} and β_m assume values in the second set. The sign here is positive since there are two interchanges of values from the first to second set.

Therefore, we have

$$\begin{bmatrix} \beta_{m-1} & \beta_m & \cdots & \beta_{2m} \\ m-1 & m & \cdots & 2m \end{bmatrix} a_{1\beta_{m+1}} a_{2\beta_{m+2}} \cdots a_{m-2\,\beta_{2m-2}} a_{\beta_{m-1}\beta_{2m-1}} a_{\beta_m \beta_{2m}}$$

$$= |a_{gh}| - \sum_r |a_{ij}| - \sum_r |a_{ik}| + \sum_r |a_{pn}| = |a_{gh}| - |a_{gh}| - |a_{gh}| + \sum_r |a_{pn}| = 0$$

where $g = 1,2,\ldots\ldots m$ $\quad h = m+1, m+2, \ldots\ldots\ldots 2m$

$l = 1,2,\ldots m-2, r, m$ $\quad j = m+1,\ldots r-1, m-1, r+1,\ldots 2m$

$i = 1,2,\ldots m-1, r$ $\quad k = m+1,\ldots r-1, m, r+1,\ldots\ldots 2m$

$p = 1,\ldots\ldots r_1, r_2$ $\quad n = m+1,\ldots r_1-1, m-1, r_1+1,\ldots$
$r_2-1, m, r_2+1\ldots 2m$

$r_1, r_2 = m+1, m+2,\ldots\ldots 2m.$

Hence $|a_{gh}| = \sum_r |a_{pn}|$.

CASE III. In an exactly similar manner we see that the expression

$$A \equiv \begin{bmatrix} \beta_{m-2} & \beta_{m-1} & \beta_m & \ldots & \beta_{2m} \\ m-2 & m-1 & m & \ldots & 2m \end{bmatrix}$$

$$a_{1\beta_{m+1}}\, a_{2\beta_{m+2}} \ldots\ldots a_{m-3\,\beta_{2m-3}}\, a_{\beta_{m-2}\,\beta_{2m-2}}\, a_{\beta_{m-1}\,\beta_{2m-1}}\, a_{\beta_m\,\beta_{2m}} = 0.$$

As β_{m-2}, β_{m-1} and β_m assume all possible values from $(m-2\ldots 2m)$ and for each set of these values, the remaining β's assume all possible arrangements of the remaining fixed values, we get a summation of determinants. Some of these may be grouped according to case I (and there will be ${}^3C_1$, such groups); others under case II (${}^3C_2$ such groups here) and finally a summation of determinants, in which β_{m-2}, β_{m-1} and β_m all three assume values out of the second set.

Making use of the results obtained under case I and case II the above expression (A) may be written

$$A = |a_{gh}| - 3|a_{gh}| + 3|a_{gh}| - \sum_r |a_{uv}| = 0.$$

Hence $|a_{gh}| = \sum_r |a_{uv}|$

where $u = 1,\ldots\ldots m-3, r_1, r_2, r_3$

$v = m+1,\ldots r_1-1, m-2, r_1+1,\ldots r_2-1, m-1, r_2+1,\ldots$
$r_3-1, m, r_3+1,\ldots 2m$

$r_1, r_2, r_3 = m+1, m+2,\ldots\ldots 2m.$

By comparing the results of cases I, II, III, a statement for the general case, where there are any number of summation symbols in the first set, may be made.

We note that the summations of determinants may be grouped so as to get $\pm k|a_{gh}| \pm \sum_r |a_{uv}|$; the signs alternating, and k, the coefficients being those of the binomial expansion. Hence, we have for the case of four summation symbols in the first set

$$|a_{gh}| - 4|a_{gh}| + 6|a_{gh}| - 4|a_{gh}| + \sum_r |a_{uv}| = 0.$$

Hence $|a_{gh}| = \sum_r |a_{uv}|$

where $u = 1,\ldots\ldots r_1, r_2, r_3, r_4$

$v = m+1,\ldots\ldots r_1-1, m-3, r_1+1,\ldots r_2-1, m-2, r_2+1,\ldots$
$r_3-1, m-1, r_3+1,\ldots r_4-1, m, r_4+1,\ldots 2m$

Therefore, if we have s summations symbols in the 1st set of values, the general theorem is

$$|a_{gh}| = \sum_r |a_{uv}|$$

where $g = 1,2,\ldots\ldots m$

$h = m+1, m+2,\ldots\ldots 2m$

$u = 1,2,\ldots\ldots m-s, r_1, r_2,\ldots r_s.$

$v = m+1, m+2,\ldots r_1-1, m-s+1, r_1+1,\ldots r_2-1, m-s+2,\ldots r_s-1, m, r_s+1,..2m$

$r_1, r_2\ldots\ldots r_s = m+1, m+2,\ldots\ldots 2m.$

The number of determinants in the summation on the right equals ${}^mC_s$. Hence we see that Kronecker's Theorem is a special case of the theorem stated here—the case where $s = 1$.

[1] We wish to express our indebtedness to Professor F. D. Murnaghan for suggesting the possibility of extending this relation, after reading Professor E. B. Stouffer's paper on this subject in these PROCEEDINGS, January, 1926.

[2] Murnaghan, F. D., *Amer. Math. Monthly*, **32**, May, 1925 (233–241). Murnaghan, F. D., *Bull. of Amer. Math. Soc.*, **31**, July, 1925 (323–329).

ON THE CLASSICAL DIRICHLET PROBLEM FOR GENERAL DOMAINS

BY OLIVER D. KELLOGG

DIVISION OF MATHEMATICS, HARVARD COLLEGE

Communicated May 17, 1926

1. *Introduction.*—We shall be concerned with the most general *domains*, or open continua, in space of three dimensions. It will be convenient to denote domains, and points in them, by capital letters, and to denote their boundaries, and points of these boundaries, by corresponding small letters. We shall consider explicitly only domains, T, whose boundaries, t, are bounded point-sets, a restriction which can be removed by inversions.

A function $F(p)$, defined on t, is said to be continuous if to every $\epsilon > 0$, there corresponds a $\delta > 0$, such that $|F(p)-F(q)| < \epsilon$ for any two points of t whose distance, pq, $< \delta$. By the classical Dirichlet problem, we mean that of determining a function, $U(P)$, harmonic in T, such that $U(P) \longrightarrow F(p)$ as $P \longrightarrow p$. If T is not bounded, the additional restriction is placed on $U(P)$ that it shall vanish at infinity like the potential of bounded charges.

The problem is not always possible. Zaremba[1] and Lebesgue[2] have

exhibited cases in which exceptional boundary points exist, at which functions having otherwise the properties of solutions cannot approach the assigned boundary values. But in any case a function $U(P)$, bounded and harmonic in T, can be associated with given continuous boundary values, $F(p)$, which will approach $F(p)$ at every regular (not exceptional) boundary point.[3] Wiener[4] derived important properties of this function $U(P)$, which we shall call the *sequence solution* of the Dirichlet problem.

We shall establish the results of Wiener in a way simpler than has heretofore been given, taking as point of departure a generalized function of Green for T, and shall obtain additional light on the sequence solution. We shall then give certain theorems on the *capacity* of point-sets, and show, in particular, that sets of 0 capacity are the most general at which harmonic functions can have removable singularities. Finally, we show that if the exceptional points of t form a set of 0 capacity, the sequence solution is the only function which is bounded and harmonic in T and which approaches $F(p)$ at all regular boundary points.[5]

2. *Green's Function.*—Let $[T_n]$ denote a nested set of *normal* domains with T as limit; i.e., a set such that the classical Dirichlet problem is possible for each T_n and any continuous boundary values, such that T_n is contained in all the following domains of the set, and such that each point of T is contained in some T_n. Then, by the reasoning of Harnack,[6] the functions of Green, $G_n(P,Q)$, for the domains, T_n, and fixed pole, Q, form a sequence which approaches a limit uniformly in any closed sub-region of T omitting Q. This limit, $G(P,Q)$, we shall call Green's function for T, leaving open the question as to whether it vanishes at all boundary points or not.

From the manner of its formation, $G(P,Q)$ is seen to have the following properties. $G(P,Q)-1/PQ$ is harmonic in T, save for a possible removable singularity at Q; if T is not bounded, $G(P,Q)$ vanishes regularly at infinity; $G(Q,P) = G(P,Q)$.

THEOREM I. *$G(P,Q)$ is independent of the particular set of nested domains, $[T_n]$, used in its definition.* For, if $[T'_n]$ denote a second set, $G(P,Q)$ dominates all the functions of Green for the domains T'_n, and hence their limit, $G'(P,Q)$, i.e., $G'(P,Q) \leqq G(P,Q)$. Similarly, $G(P,Q) \leqq G'(P,Q)$, and the theorem follows.

If T is a normal domain, $G(P,Q) \longrightarrow 0$ as $P \longrightarrow p$, p being any boundary point. The converse was first proved by Lichtenstein.[6] If T is not normal, there will be boundary points at which $G(P,Q)$ does not approach 0. We define as *regular* and *exceptional* boundary points, those at which $G(P,Q)$ does, or does not, approach 0, respectively.

THEOREM II. *The definition of regular and exceptional points is independent of the position of the pole, Q.* Let Q and Q' be two points of T, and let R be a region consisting of a domain containing Q and Q', together

with its boundary, r, and lying in T. Let m be the positive minimum of $G(P,Q')$ on r, and M the maximum of $1/PQ$ on r. As $G_n(P,Q) \leqq 1/PQ$, it follows that $G_n(P,Q) \leqq (M/m)G(P,Q')$ on r, and since this inequality holds on t_n, the boundary of T_n, it holds also throughout $T_n - R$. Hence the limiting form of this inequality, $G(P,Q) \leqq (M/m)G(P,Q')$ holds in $T-R$, and $G(P,Q)$ must approach 0 at any boundary point at which $G(P,Q')$ does.

The above reasoning, with the inequality given, leads at once to

LEMMA I. *Let Q be confined to a closed sub-region, R, of T. Then*

(*a*) *as P approaches a regular boundary point, $G(P,Q) \longrightarrow 0$ uniformly as to Q,* and

(*b*) *if p is exceptional, but such that for fixed Q' a sequence $[P_k]$ with p as limit point exists for which $G(P_k,Q') \longrightarrow 0$, $G(P_k,Q) \longrightarrow 0$, uniformly as to Q, as $P \longrightarrow p$.*[7]

3. *The Sequence Solution.*—The sequence solution is derived from any given continuous boundary values, $F(p)$, just as was Green's function for the boundary values $1/PQ$. To $F(p)$ there correspond infinitely many functions, continuous in $T + t$, and coinciding with $F(p)$ on t. Let $F(P)$ be such a function. Then, to each of an infinite sequence of nested normal domains, $[T_n]$, there corresponds a function, $u_n(P)$, harmonic in T_n, and coinciding with $F(P)$ on t_n. Wiener's theorem is that the sequence $[u_n(P)]$ approaches a limit, $U(P)$, uniformly in any closed sub-region of T, and that $U(P)$ is independent of the particular extension of $F(p)$ to the points of T, and of the set $[T_n]$ employed. In calling $U(P)$ the sequence solution of the Dirichlet problem for T and $F(p)$, we are, of course, using the word solution in an extended sense, for a strict solution is not always possible.

We turn now to a study of the sequence solution, starting with the case in which $F(p)$ are the boundary values of a polynomial, $F(P)$, in the cartesian coördinates of the point P of T. We also assume first that T is bounded. The Laplacian of $F(P)$ is again a polynomial,

$$\nabla^2 F(P) = f(P). \tag{1}$$

We form the function

$$\phi(P) = \lim_{n \longrightarrow \infty} \phi_n(P), \qquad \phi_n(P) = \frac{1}{4\pi} \iiint_{T_n} f(Q)G(P,Q)dV_Q. \tag{2}$$

In case T_n fails to have Jordan content, the integral above is to be understood as the limit of the integral over a set of nested domains with content, and with T_n as limit. As $f(P)$ is a polynominal, and as $G(P,Q) \leqq 1/PQ$, the sequence $[\phi_n(P)]$ is seen to be uniformly convergent in any closed sub-region of T, and its limit, $\phi(P)$ to be independent of the particular set $[T_n]$ used.

THEOREM III. *The function* $U(P) = F(P) + \phi(P)$ *is bounded and harmonic in* T, *and approaches the boundary values* $F(p)$ *at every regular boundary point.*

$U(P)$ is obviously bounded. If, in the integral for $\phi_n(P)$, $G(P,Q)$ be written $1/PQ + g(P,Q)$, the Laplacian of the first integral thus arising is well known to be $-4\pi f(P)$, while that of the second is 0. Hence $\nabla^2\phi_n(P) = -f(P)$, and so $\nabla^2\phi(P) = -f(P)$ because the sequence $[\phi_{n+m}(P) - \phi_n(P)]$ is, for fixed n, uniformly convergent in T_n, and its terms are harmonic in T_n. Thus, as $\nabla^2\phi(P) = -f(P)$, and $\nabla^2 F(P) = f(P)$, $\nabla^2 U(P) = 0$.

It remains to show that $U(P) \longrightarrow F(p)$, that is, that $\phi(P) \longrightarrow 0$, as $P \longrightarrow p$, a regular boundary point. Since $\phi(P)$ is independent of $[T_n]$, we may assume the boundary of each T_n to be interior to all following domains of the set. Then, for any fixed n, and any $\epsilon > 0$, there is a $\delta > 0$, such that when $Pp < \delta$, $G(P,Q) < \epsilon$ for all Q in T_n. Hence, by lemma I, if M is the upper limit of $f(P)$ in T, and if B is the volume of a cube containing T,

$$|\phi_n(P)| \leqq \frac{MB}{4\pi}\epsilon \tag{3}$$

for $Pp < \delta$. On the other hand, for positive integral m,

$$|\phi_{n+m}(P) - \phi_n(P)| \leqq \frac{M}{4\pi}\iiint_{T_{n+m}-T_n} G(P,Q)dV_Q \leqq \frac{M}{4\pi}\iiint_{T_{n+m}-T_n}\frac{1}{PQ}\,dV_Q,$$

and the last integral, extended over a region of given volume is least when the region is a sphere with P as center. Hence

$$|\phi_{n+m}(P) - \phi_n(P)| \leqq \frac{M}{8\pi}\left(\frac{3V}{4\pi}\right)^{2/3}, \tag{4}$$

where V is the inner content of T less the volume of T_n. For large n, V, and therefore $|\phi_{n+m}(P) - \phi_n(P)|$, is arbitrarily small, independently of m and P. Then, with n fixed, the right-hand member of (3) may be made arbitrarily small by sufficiently restricting Pp, and hence the same is true for $\phi_{n+m}(P)$, independently of m. The same is, therefore, true of $\phi(P)$, that is, $\phi(P) \longrightarrow 0$ as $P \longrightarrow p$.

If p is an exceptional boundary point, but one for which a sequence $[P_k]$, as described in lemma I (b), exists, then by the same reasoning, $\phi(P_k) \longrightarrow 0$ as $P_k \longrightarrow p$. Thus the sequence solution approaches the given boundary values on certain sequences of points even at exceptional points, provided these sequences of points exist. And they do exist, save at certain boundary points at which possible discontinuities of the solution are removable (p. 404).

We have assumed, in the proof of theorem III, that T was bounded. If this is not the case, but t is bounded, say, by a sphere with center O and

radius R, we proceed as follows, still assuming that $F(p)$ are the boundary values of a polynomial, $F_1(P)$. We form a function, $H(P)$, as follows: when $r = OP \leqq R$, $H(P) = 1$; when $r \geqq 2R$, $H(P) = 0$, and when $R \leqq r \leqq 2R$, $H(P)$ is a polynomial of degree 5 in r, so chosen that $H(P)$ and its first two derivatives with respect to r are continuous. The continuous extension of $F(p)$ to the points of T is then $F(P) = F_1(P)H(P)$. Then $\nabla^2 F(P)$ is bounded, and identically 0 for $r \geqq 2R$. The integrals are then again taken over bounded domains.

THEOREM IV. *The function $U(P)$ of the preceding theorem is the sequence solution of the Dirichlet problem for T and $F(p)$.* This will follow at once if it is clear that $G_n(P,Q)$ may replace $G(P,Q)$ in the formula (2) without changing the limit, for then $\phi_n(P) + F(P)$ is exactly the sequence function $u_n(P)$. The replacement in question, however, does not affect the limit, because of the uniform convergence of $[G_n(P,Q)]$ to $G(P,Q)$ in any closed sub-region of T, and of the fact that $G_n(P,Q) \leqq G(P,Q) \leqq 1/PQ$. Thus the sequence solution exists, and is independent of $[T_n]$, and approaches the given boundary values in the ways described above, all on the assumption of polynomial boundary values—a restriction which we now remove.

Let $F(p)$ be any continuous function, and let $F(P)$ be a function, continuous in $T + t$, and coinciding with $F(p)$ on t. Let $F'(P)$ be a polynomial which differs from $F(P)$ in $T + t$ by less than ϵ, and let $u_n(P)$ and $u'_n(P)$ be the sequence functions for the same set $[T_n]$ and with the values $F(P)$ and $F'(P)$ on t_n. Then $u'_n(P) - \epsilon \leqq u_n(P) \leqq u'_n(P) + \epsilon$ in T_n. As $[u'_n(P)]$ approaches a limit, $U'(P)$, uniformly in any closed sub-region, R, of T, for all great enough n, and all P in R, $U'(P) - 2\epsilon \leqq u_n(P) \leqq U'(P) + 2\epsilon$. And as there is a relation of this form for every positive ϵ, $[u_n(P)]$ approaches a limit, $U(P)$, uniformly in R. Furthermore, the limiting forms of the last inequalities show that the upper and lower limits of $U(P)$ at each regular boundary point lie between $F(p) - 3\epsilon$ and $F(p) + 3\epsilon$. But as these limits are independent of ϵ they coincide and $U(p) \longrightarrow F(p)$. The situation is similar in the case of sequence approach at certain exceptional points.

Finally, the sequence solution is independent of the continuous extension of $F(p)$ to the points of T. If $F(P)$ and $K(P)$ are two such extensions, and $[u_n(P)]$ and $[v_n(P)]$ the corresponding sequences, the sequence $[u_n(P) - v_n(P)]$ whose terms coincide with $F(P) - K(P)$ on t_n approaches 0 uniformly, since $F(P) - K(P)$ approaches 0 uniformly at the points of t. Hence the sequence solutions $U(P)$ and $V(P)$ corresponding to $F(P)$ and $K(P)$ coincide.

If, in the above use of approximating polynomials, T is not bounded, we may take any continuous extension, $F(P)$, of the given boundary values, and, using the function $H(P)$, previously established, approximate to $F(P)H(P)$ by a polynomial in the portion of T within a sphere about O

of radius $2R$. Instead of this polynomial, we then use its product by $H(P)$ for $r < 2R$ and 0 for $r \geqq 2R$.

4. *The Capacity of Point-Sets.*—It is possible, as Wiener[8] has shown, to extend the notion of the capacity of a conductor to any bounded point-set. We shall give his definition, in slightly modified form, and give some properties and applications of capacity.

Let B be any bounded point-set, and let B' be the set consisting of B and its limit points. B' may bound a number of domains, but a part of B', which we denote by t, will bound a domain, T, extending to infinity. Consider the sequence solution, $v(P)$, of the Dirichlet problem for T and the boundary values 1 on t. We call $v(P)$ the *conductor potential* of the set B. The *capacity* of B is then the total charge of which $v(P)$ is the potential, or the Gauss integral, $\frac{1}{4\pi}\int\int \frac{\partial v}{\partial n}\,dS$, extended over any smooth surface containing B in its interior, n being the normal to this surface, directed inward.

THEOREM V. *If A is a part of B, the capacity of A is less than, or equal to, that of B.* The continuous extension of the boundary values may be taken as 1 itself, and the nested regions, $[S_n]$ and $[T_n]$, used in constructing the conductor potentials $w(P)$ and $v(P)$ of A and B, respectively, may be so chosen that S_n is always a part of T_n. Then, on t_n, $v_n(P) = 1$, while $w_n(P) \leqq 1$. Hence $v(P)-w(P)$ is never negative. If $v(P)-w(P)$ vanishes at any point of T, it vanishes identically, and the capacities of A and B are then equal. Otherwise, let R denote an equipotential surface $v(P)-w(P) = C$, C being small enough to insure this surface being the smooth boundary of a simply connected domain. On R, $\frac{\partial}{\partial n}[v(P)-w(P)]$ cannot be negative, since if it were, $v(P)-w(P)$ would exceed C at some point outside R. Hence the Gauss integral over R cannot be negative, and the capacity of A is, therefore, less than, or equal to, that of B.

THEOREM VI. *Let T denote the infinite domain bounded by a part of B and its limit points, and $[T_n]$ a nested set of normal domains with T as limit. Then the capacities of the boundaries, t_n, of T_n, converge, decreasing monotonely, to the capacity of B.* This follows from theorem V, and the uniform convergence in any closed sub-region of T of the sequence $[v_n(P)]$ defining the conductor potential of B, together with the uniform convergence of the sequence of normal derivatives of $v_n(P)$ to the corresponding derivatives of $v(P)$.

LEMMA II. *If B is a bounded set of capacity $c>0$, the value of the conductor potential of B at a point, P, of T, is subject to the inequalities*

$$\frac{c}{p'P} \leqq v(P) \leqq \frac{c}{p''P},$$

where p' and p'' are the points of B' (B and its limit points) farthest from, and nearest to P, respectively. Because of theorem VI, and the independence of the sequence solution, $v(P)$, of the nested set, $[T_n]$, it will suffice if the lemma is proved on the assumption that B consists of a finite number of surfaces with continuous principal curvatures. Then $v(P)$ is the potential of a continuous distribution of a charge, c, on B, and $v(P)$ is decreased by concentrating the charge at the point of B farthest from P, and increased by concentrating it at the nearest point.

COROLLARY. *If B is a bounded set of capacity 0, its conductor potential vanishes identically in T.* For B is enclosable in the boundary, t_n, of a normal domain, T_n, the capacity of t_n being arbitrarily small. Thus the conductor potential of B is dominated at P by an arbitrarily small number, that is, it is 0.

5. *Removable Singularities of Harmonic Functions.*—A function which is bounded and harmonic in the neighborhood of a point, or of an arc of a smooth curve, can be so defined at such points as to be harmonic there also. That is, such points or arcs can be the seat only of removable singularities. A complete generalization of this fact is possible by means of the notion of capacity.

THEOREM VII. *Let T be any domain, and let B be any part of its boundary with the following properties: (a) the set $T' = T + B$ is still a domain, and (b) the capacity of B is 0. Then any bounded function, harmonic in T, may be so defined on B as to be harmonic in T'.*[9] To show this, we first establish the following lemma.

LEMMA III. *Let σ be the surface of a sphere, and $F(p)$ a function of p on σ, bounded, and continuous except at the points of a set, s, of capacity 0. Then there exists a function, $U(P)$, harmonic and bounded within σ, and approaching the boundary values $F(p)$ at all points where this function is continuous.* We assume $0 \leqq F(p) \leqq 1$. The proof can then be extended to the more general case by a linear transformation. We assume also that s is closed. Its capacity is not affected by adding its limit points. By theorem VI, for every positive integer n, there exists a finite number of smooth surfaces containing s in their interior, and of capacity less than $1/n$. Within these surfaces, a finite number of smooth curves can be drawn on σ, bounding a domain, s_n, on σ, which contains s, and whose capacity is less than $1/n$, by theorem V. Furthermore, these domains s_n can be chosen so that each includes all the following.

If we define $F'(p) = 0$ on s_n, and $F'(p) = F(p)$ elsewhere on σ, the discontinuities of $F'(p)$ are limited to a finite number of smooth curves, so that Poisson's integral defines a function $u_n(P)$, harmonic within σ, and approaching the boundary values $F'(p)$ save on the curves bounding s_n. The sequence $[u_n(P)]$ is never decreasing, and so approaches a harmonic limit, $U(P)$, uniformly in any closed region within σ. The bounds of

$U(P)$ are clearly 0 and 1. It remains to show that $U(P) \longrightarrow F(p)$ as $P \longrightarrow p$, a point of continuity of $F(p)$. Let R be any closed region of space containing no point of s. Then, given $\epsilon > 0$, n can be taken so large that the conductor potential, $v_n(P)$, of s_n, is less than ϵ in R, by theorem VI. But on $\sigma - s_n$, $u_{n+m}(P) - u_n(P) = 0$, for any positive integer m, and this difference never exceeds 1. Hence $-v_n(P) \leqq u_{n+m}(P) - u_n(P) \leqq v_n(P)$ on σ, and hence also in its interior.[10] Hence the sequence $[u_n(P)]$ converges uniformly in the closed region consisting of the points of R within or on σ, and as the terms of the sequence have the required boundary values there, so does their limit.

Returning now to the theorem, let Q be any point of B. Since T' is a domain, Q is the center of some sphere, σ, lying entirely in T'. Let $V(P)$ be any function, bounded and harmonic in T. Then its values on σ define a function $F(p)$ satisfying the conditions of the lemma, by theorem V. If $U(P)$ be the function, harmonic in σ, determined by these values $F(p)$ in accordance with the lemma, $V(P) - U(P)$ is bounded within and on σ, is harmonic within σ except possibly at the points of B, and has the boundary values 0 on σ except at the points of B. If the bound of the absolute value of $V(P) - U(P)$ is M, this difference, and its negative, is dominated by the conductor potential of any set of surfaces, t_n, enclosing B. Hence, as the capacity of B is 0, $V(P) - U(P) = 0$ throughout the interior of σ except on B. Hence if $V(P)$ is defined on the portion of B within σ as equal to $U(P)$, it will be harmonic within σ. Thus $V(P)$ may be so defined at all points of B as to be harmonic in T'.

THEOREM VIII. *Conversely, if B has the property (a) of theorem VII, and if any function bounded and harmonic in T can have only removable singularities at the points of B, then B has capacity 0.* For the conductor potential of B is bounded and harmonic in T, and so has only removable singularities. When properly defined on B, it is harmonic everywhere, and vanishes at infinity, and so vanishes identically. Hence the capacity of B is 0.

From theorem VII, it follows that boundary sets of the type B can have no influence on the classical Dirichlet problem. They are called *improper sets* by Bouligand, who defines them as sets of points at which $\underline{lim}\, G(P,Q) > 0$. It is not obvious that the concepts are equivalent, so we give a proof.

LEMMA IV. *The conductor potential of any set of positive capacity has 1 as upper limit.* For if its upper limit were $M < 1$, then, $v(P)$ denoting this conductor potential, $v(P)/M$ would be dominated by all the terms of the sequence $[v_n(P)]$ whose limit defines $v(P)$, and we should have $v(P) \geqq v(P)/M$, a contradiction.

Now let b be the set of points at which $\underline{lim}\, G(P,Q) > 0$. The set $T' = T + b$ is an open continuum, since the points of b are limit points of T, and since if a point of b were not an interior point of T', it would have to

be the limit of a set of points at which $\underline{lim}\ G(P,Q) = 0$ and so could not itself be a point at which $\underline{lim}\ G(P,Q) > 0$. We wish to prove that b has capacity 0. Let A be a point of b, and σ a sphere with center at A and lying in T'. Then, within and on σ, $G(P,Q)$ has a positive lower bound, m. Otherwise there would be in this closed region a point at which $\underline{lim}\ G(P,Q) = 0$, i.e., a point not in T'. Let s denote the set of points of b in σ, with their limit points. Then s has capacity 0. Otherwise, the conductor potential of s would have the upper limit 1. Now on a sphere σ', of radius r and center Q, lying in T, this conductor potential, $v(P)$, has a maximum, $M < 1$, so that $\frac{1}{r}\left(\frac{1-v(P)}{1-M}\right)$ is a function harmonic in T, greater than, or equal to $1/r$ on σ', and never negative. It therefore dominates the sequence functions $G_n(P,Q)$ in $T-\sigma'$, and hence also $G(P,Q)$. But it has the lower limit 0 in σ, whereas that of $G(P,Q)$ is $m > 0$. Thus the assumption that the capacity of s is positive has led to a contradiction.

It is now possible to remove from the boundary of T a portion of b, contained in s but containing all the points of b in a sphere about A, so as still to leave an open continuum in place of T. The set removed, being a part of s, has capacity 0. Theorem VII then shows that any function, harmonic and bounded in T may be so defined on this portion of b as to be harmonic at A. But as A was any point of b, the harmonic function may be defined at all points of b so as to be harmonic in $T' = T + b$. Hence by theorem VIII, b is of capacity 0, and so of the type of the set B.

Conversely, if P is any point of a set of type B, $\underline{lim}\ G(P,Q) > 0$ at P. For, after the definition of $G(P,Q)$ on B has been properly made, $G(P,Q)$ is harmonic at P, and so continuous and positive.

We close with a uniqueness theorem.

THEOREM IX. *If the exceptional boundary points of T form a set of capacity 0, there is one and only one function, $U(P)$, bounded and harmonic in T, which approaches preassigned continuous boundary values, $F(p)$, at every regular boundary point.* One such function is the sequence solution. If there were two, their difference, $V(P)$, would be bounded and harmonic in T, and approach O at every regular boundary point. If M is a bound for $|V(P)|$, $-V(P)/M$ and $V(P)/M$ are dominated by the conductor potential of any set including the exceptional points of t, and so, by a now familiar argument, vanish.[11]

[1] Zaremba, *Acta Math.*, **34** (1911), p. 310.

[2] Lebesgue, *C. R. des Sce. de la Soc. Math. de France* (1913), p. 17.

[3] Kellogg, *Proc. Amer. Acad.*, **58**, No. 14 (1923), p. 528.

[4] "Certain Notions in Potential Theory," *Jour. of Math. and Phys. of the Mass. Inst. of Tech.*, **3** (1924), p. 25.

[5] Results closely parallel to a number of those to be given have recently been published by Bouligand in two stimulating memoirs, *Mémorial des sci. math.*, Paris, 1926, and *Annales de la Soc. Polonaise de Math.* (1925), pp. 59–112. Of these, some had

been indicated to the present writer in correspondence, early in December, 1925. But they had been arrived at independently, in connection with his symposium on the Dirichlet problem at the Christmas meeting of the American Mathematical Society, 1925, with the exception of the results on improper sets, where again, the methods differ.

[6] *Grundlagen d. Theorie des log. Potentials, Leipzig*, 1887, p. 118, ff.

[7] See Bouligand, *Ann. Polonaise*, p. 84.

[8] See Wiener, *l. c.*, p. 26.

[9] See Bouligand, *Ann. Polonaise*, p. 104.

[10] The proof requires the fact that $v_n(P)$ approaches *1* at all points of s_n. This follows from the fact that $v_n(P)$ dominates the conductor potentials of certain spherical triangles inscribed s_n, and the fact that spherical triangles bound normal domains. See Wiener, "The Dirichlet Problem," *Jour. of Math. and Phys. of the Mass. Inst. Tech.*, **3** (1924), p. 104, with the addendum that parallel projections onto planes of point sets never have greater capacities than the sets projected.

[11] It is likely that the exceptional points of every domain form a set of capacity 0, in which case the above theorem would give a certain finality to the sequence solution. The desired result would follow from a lemma: Every closed set of positive capacity contains regular points. But the difficulties in the proof of this lemma have not been overcome.

THE EVIDENCE FOR CHANGES IN THE RATE OF ROTATION OF THE EARTH AND THEIR GEOPHYSICAL CONSEQUENCES

BY ERNEST W. BROWN

YALE OBSERVATORY, NEW HAVEN

Read before the Academy April 26, 1926

This paper is a summary of one which will shortly appear in the *Transactions of the Yale Observatory* and, in fact, constitutes the Introduction to it. The work was undertaken as a fresh attempt to see whether further light could be thrown on the question of the constancy of the earth's rate of rotation, that is, of our measure of time. The chief evidence for changes of this character is derived from observations of the moon. I am not here mainly concerned with the secular changes due to tidal friction, but with the considerable fluctuations which are exhibited in the difference between the observed and calculated longitude of the moon when all known causes of variation have been eliminated. The numerous investigations into ancient eclipses culminating in the results of Fotheringham and the work of Taylor and Jeffries on Tidal Friction in shallow seas are in substantial agreement as to the amount of the frictional effect so that it may be regarded as known. The tabular place of the moon due to gravitational theory is therefore first corrected for this effect, the fluctuations referred to being the differences between this corrected theory and observation.

Since all classes of observations of the moon in all places show these deviations, there is practically no doubt that they are real, that is, that they are not due to errors of observation, the latter being, in general, small compared with the magnitude of the deviations. Nevertheless, the differences between the results obtained from various methods of observation and at different places, show some peculiarities which are brought out below for examination in order to clear the ground.

The main issue is whether these fluctuations are due to forces, not hitherto recognized, acting on the moon so as to alter its motion, or whether they are changes in the rate of rotation of the earth. Either hypothesis will account for the observed data so far as the moon is concerned. Distinction between them can only be made by observations on other bodies, the most important of which, on account of the material available, is the sun. The conclusion reached here is that the evidence favors the latter hypothesis. This being granted, a discussion is made of the possible sources of changes in the rate of rotation of the earth with such numerical evidence as could be obtained.

The following brief account of the history of the discovery and of the attempts at explanation of these fluctuations will assist in forming an estimate of the degree of reliability of the evidence.

Newcomb first came across the traces of the fluctuations in Part I of his Researches on the motion of the moon,[1] in which he discussed various observations made before the year 1750. By 1903, in a brief paper,[2] he was able to show that they had a real existence, and six years later he gave a summary[3] of the second part of his Researches[4] which finally appeared in 1912. He there showed that the chief fluctuation appeared to be well represented by a harmonic term with a period of some 270 years and an amplitude of about 14″. Superimposed on this were minor fluctuations of some three or four seconds amplitude but irregular in period. Newcomb discussed the origin of these terms and, in view of the attempts to determine the complete gravitational solution of the lunar problem, leaned strongly towards the hypothesis that they represented changes in the earth's rate of rotation. He mentions that he tried to find confirming results from the transits of Mercury but that he was not successful.

The next extensive investigation was that of Cowell which was published in a series of papers in the *Monthly Notices* from 1903–5. In these he analyzed the Greenwich meridian observations of the moon from 1750 to 1900. Newcomb's results were confirmed and Cowell attempted an analysis of the fluctuations into two or three harmonic terms. Shortly after, in 1913, my theory of the gravitational problem of the moon's motion being complete, and the new tables being under way, I took Cowell's results and corrected them so as to make them correspond as far as necessary with the new theory, mainly in order to get the best possible

values of the principal constants. Incidentally, a comparison with Newcomb's results was made and a detailed determination of the minor fluctuations from both sources resulted. Within the degree of accuracy expected, there was good agreement in general except in the interval between 1815 and 1845 where there were divergences which I was unable to trace down.

In 1923, Dyson and Crommelin[5] took the results of Cowell and myself, carried them to the end of 1922, corrected them in certain particulars and made a full analysis of the minor fluctuations. Shortly after H. Spencer Jones[6] published his analysis of the Cape occultations from 1880–1922, mainly with a view of testing the adopted constants of the theory, but also obtaining the run of the fluctuations during that period. A further short but useful series is that deduced by Russell from photographs taken at Harvard College Observatory during the years 1911–17.[7]

In the meantime attempts were made to see if these outstanding fluctuations could be explained as deviations of the moon's motion. The detailed theory of the sun's action on the moon and searching investigations into planetary and other gravitational effects made it finally clear that all such sources must be excluded. In 1910 I examined some hypotheses designed to explain them, postulating forces other than gravitational, but with little success. The answer to a question sometimes asked as to whether the Einstein effect might be responsible is clearly in the negative. One reason is that it is too small. A second, which applies to many similar hypotheses is that small disturbing forces which are functions only of the configuration of the system will not in general produce new periods, but only combinations of periods already present, and the periods which might be present in the fluctuations do not seem to have any such property.

It appears therefore that we must fall back on the hypothesis that the fluctuations are due to changes in the earth's rate of rotation. In 1914, I was able to throw some light on the question by exhibiting evidence of similarity between the minor fluctuations exhibited by the moon's motion and the deviations of the sun and Mercury from their theoretical orbits during the period 1750 to 1900.[8] The following year Larmor[9] examined the hypothesis that changes in the earth's rate of rotation might be caused by vertical oscillations of limited areas but found that the amounts to be postulated were not in accordance with geophysical evidence. The same year, H. Glauert[10] showed that a marked similarity between the fluctuations for 1865–1915 existed in the case of material derived from Greenwich observations of the sun, Venus and Mercury, and he also showed that their order of magnitude was not very different from that to be expected on the hypothesis of changes in the earth's rate of rotation. In 1916, Ross[11] showed that Mars during the nineteenth century exhibited fluctuations very similar to those of the sun. In 1925, Innes[12] gave additional evidence from the transits of Mercury and the solar deviations.

Thus evidence has gradually accumulated which seems to justify the hypothesis. In view of the new material on the moon's motion which has been assembled during the past few years, it seemed worthwhile to examine the material once again in a collected form. This first demanded a detailed comparison of the observations of the moon with theory, which, as already stated, was in large measure already available, and with this an intercomparison of the various series of observations to determine, if possible, the sources of systematic differences. To test the hypothesis, the fluctuations should be visible in the motion of the sun, but only to one-thirteenth of the amount. It is necessary to combine with them the secular acceleration of the sun discovered by Cowell[13] in 1905 and later verified by Fotheringham[14] from a discussion of ancient and modern observations; the results were in large measure confirmed by Jeffries[15] in his calculations of the tidal friction in shallow seas. This demanded an examination of the sun's deviations from its tabular positions and here again most of the material had been made available by Newcomb and Ross. While in the previous attempts the effect of the minor fluctuations of the secular acceleration had been examined, the whole effect, namely, the total fluctuations as determined from the moon added to the sun's secular acceleration, do not seem to have been applied to the observations of the sun.

The results obtained from the present examination may be briefly summarized as follows:

1. Observations of the sun since 1750 are consistent with the hypothesis that the observed fluctuations of the moon are due to variations in the earth's rate of rotation, when these are combined with the known effects of tidal friction. In particular, the hypothesis accounts for the principal part of the sudden change in the deviations of the sun from its tables which began in 1900 and which is at present greater than 1″. The previously obtained correlations between these differences and those deduced from observations of the planets are also consistent with the hypothesis but are partly to be explained by common systematic errors of observation.

2. Granting the hypothesis, arguments resulting from astronomical data, mechanical laws and geophysical evidence point to the conclusion that such variations can only arise from vertical oscillations of the whole crust of the earth, the source of these oscillations being due to oscillatory changes taking place in the isostatic layer or below it. The maximum change of the average radius needed by the astronomical data is between five inches and twelve feet, according to the hypothesis made as to the depth of the source.

3. Evidence in favor of the following correlations appears to be sufficiently good for an explanation of them to be needed.

The differences between the Greenwich meridian observations of the moon and the occultations as reduced by Newcomb are affected with a

fluctuating error which has the same principal maxima and minima as the minor fluctuations themselves, but the ratio of the amplitude of the former to that of the latter diminishes from about unity at the beginning of the nineteenth century to 1/8 at the present time.

4. A fluctuation with a shorter period in these same differences appears to be related to the frequency of occurrence of British earthquakes.

5. During the whole run of the Washington observations of the sun (1846 to date), the differences between them and the tabular values follow the same major fluctuations as the differences between the Greenwich observations of the sun and the tables. The change appears to have a period of about forty years.

6. The large deviations of the Greenwich observations of the sun from the tabular values between 1805 and 1825, taken in conjunction with the similar large deviations of the Greenwich meridian observations of the moon from the occultations, point to the conclusion that the meridian observations are defective in this interval and that it is unsafe to make use of them with full weight for the discussion of tabular differences, unless the source of the apparent errors can be found and the errors, if existent, corrected. The indications are that for discussion of variations of the earth's rate of rotation, the occultations rather than the meridian observations of the moon should be used in this period. It follows that harmonic analyses based on the meridian observations before 1850 are of doubtful reality and that the minor fluctuations are much more irregular in character than had previously been suspected.

7. It is doubtful whether attempts at harmonic analysis of the fluctuations will give any clues as to their origin.

8. A somewhat doubtful correlation is that between the Greenwich (and Washington) tabular differences for the sun and the differences between the Greenwich meridian observations of the moon and the occultations since 1870. Another, which may be accidental, between the same series appears before 1840 but it requires a lag of ten years in the former in order to make it correspond with the latter.

In the discussion on the forces which can change the earth's angular velocity, where I have attempted to prove that the only hypothesis which can account for the facts is an oscillatory change in the earth's mean radius. two calculations have been made. In one the earth is supposed to expand and contract throughout its whole mass uniformly. In the other, all the expansion and contraction is supposed to be due to changes of pressure in a layer 50 miles deep, changing the height of the crust above it. These are set down merely as outside limits which are in accordance with the knowledge we have at present concerning the earth. The limits can possibly be narrowed, but I am not concerned here with attempts to do so nor have I tried to make any suggestion as to the physical or chemical mechanism by

which such expansions or contractions can be produced. As I view the problem, the argument depends on a chain of reasoning which appears to require the existence of some such force well below the crust of the earth. The strength of the chain depends on the strength of its weakest link. The chief steps are as follows:

1. The astronomical data. There can be little doubt that the principal part of the apparent variations in the moon's motion are real and not due to errors of observation.

2. The evidence that the observed variations are variations in the rate of rotation of the earth and not in the rates of revolution of the moon, sun and planets about the earth. The greater part of the paper is devoted to setting forth the evidence for this, involving attempts to distinguish between systematic errors of observation and real differences between gravitational theory and observation in the motions of these bodies, especially in that of the sun.

3. The arguments by which sources for these variations are traced to forces well below the outer crust of the earth. The evidence is largely negative, that is, it consists of attempts to show that, unless matter acts on matter in some way at present quite unknown, the forces cannot proceed from any other source.

4. Assuming the last result as provisionally established, the collateral evidence which can be brought forward to support it in the form of correlations between geophysical and the astronomical data, and simplified explanations of geodynamical questions.

An oscillation with a period between 200 and 300 years in the frequency of Chinese earthquakes brought forward by H. H. Turner[16] in 1920, and compared by him with the principal fluctuation in the moon's motion may be significant in this connection.

As the main argument must ultimately depend on the numerical data, the latter have been set out in detail as far as possible with full references to their sources, and with an exact account of what changes or corrections have been made. The treatment of these data has necessarily been such as to bring out as prominently as possible the facts needed for the argument. That it also brings out other results for which explanation is not forthcoming is inevitable. The question as to whether the latter are relevant to the main issue is discussed in some detail.

While the discussion of the effect of changes in the external radius of the earth would lead to the expectation of correlation between geophysical and astronomical data, the mechanism which is responsible for a correlation between the errors (if existent) of the Greenwich observations and the frequency of certain seismic data is not apparent. Owing to the lack of uniformity of the earth's crust at least down to the isostatic layer of compensation, we may expect differential vertical motions of the crust super-

imposed on the main vertical oscillation common to the whole earth. The magnitude of that oscillation exhibited by the moon which is correlated below with the frequency of occurrence of British earthquakes seems to be rather greater than can be attributed to changes of the vertical at Greenwich. Such an explanation would seem to demand corresponding changes in the sun's motion, and there is not much, if any, evidence of them. If, however, as seems likely, the effect on the latter is only $^1/_{13}$ of that on the former their absence would be explained. The well-established correlation of the differences between the Greenwich meridian observations and the occultations, with the differences between the observations and theory would seem to result from the same cause, but it is difficult to suggest any reasonable hypothesis to account for the manner in which it is brought about.

Attempts were made to connect these differences with those obtained by Sampson[17] in the time determinations at different observatories. His proof that, at least at Edinburgh,[18] they were connected with level errors seemed to hold out some promise and I made attempts at correlation between his results and those simultaneously furnished by observations of the moon and sun, but obtained nothing of sufficient significance to reproduce here.

A theory by Prof. Joly, reported in the Observatory for 1926, February, also postulates vertical oscillations of the whole crust.[19] The basis appears to be its value in explaining the phenomena of dynamic geology. He believes that the cause is to be found in the thermal effects of radium acting in a substratum of basalt at a depth of some 30 kilometers and that it is measured in time by geological periods. This basalt was alternately liquid and solid; at the present time it is solid. In this paper I deduce from astronomical data, secured during the past two centuries, oscillations measured by years and by one or two centuries, and place the origin at least at a depth of 80 or 90 kilometers, but make no hypothesis as to their cause. The nature of the oscillation which he deduces is therefore quite different from that of this paper.

[1] *Wash. Obs. App.*, **2**, 1875.
[2] *M. N.*, vol. 63, p. 316.
[3] *M. N.*, vol. 69, p. 164.
[4] *Amer. Eph. Papers*, vol. 9, pt. 1.
[5] *M. N.*, vol. 83, p. 359, and App. G. of the Greenwich Observations for 1920.
[6] *M. N.*, vol. 85, p. 21. *Cape Obs. Annals*, vol. 8, pt. 8, 1925.
[7] *A. J.*, vol. 36, p. 153.
[8] *Brit. Assn. Report*, **1914**, p. 311.
[9] *M. N.*, vol. 75, p. 211.
[10] *M. N.*, vol. 75, p. 489.
[11] *A. N.*, vol. 225, p. 109.
[12] *A. J.*, vol. 29, p. 152.
[13] *M. N.*, vol. 66, p. 3.

[14] *M. N.*, vol. 81, p. 104.
[15] *Trans. R. S.*, vol. 221A, p. 239.
[16] *M. N.*, vol. 80, p. 807.
[17] *M. N.*, vol. 81, p. 89.
[18] *M. N.*, vol. 85, p. 560.
[19] A summary of his book, *The Surface of the Earth.*

THE REDDISH BLUE ARCS AND THE REDDISH BLUE GLOW OF THE RETINA: SEEING YOUR OWN NERVE CURRENTS THROUGH BIOLUMINESCENCE

BY CHRISTINE LADD-FRANKLIN

COLUMBIA UNIVERSITY

Read before the Academy April 27, 1926

A simple band of bright red light thrown upon a screen in a dark room gives rise to a very curious phenomenon—discovered in the first instance by Purkinje.[1] What you see on the screen is not only the red band but also, projecting out from it on both sides, big slightly reddish blue arcs. They are not of the color of the visual purple, which is a slightly bluish red. If one considers the shape and the angular size of these arcs, it is perfectly plain that what one sees, as an entoptic phenomenon, is certain fibres of the optic nerve which lie on the surface of the retina and which proceed to their exit point, the papilla. But why are they visible? The explanation that has been given of this phenomenon hitherto, by Gertz, by Troland and by others, is that the nerve current by which one sees the red band gives rise to a secondary induced current in adjoining nerve fibres. Such a current as this, however, would not be provided with the right "place coefficients"—it would not enable one to see the stimulated fibres in the place where they lie. Let me put this hypothetical question: Suppose that one could lay open an optic nerve fibre all the way from the retina to the occipital lobe of the cortex, and suppose that one were to pinch it in half a dozen different places. The pinching of this nerve fibre would give one a light sensation (either chromatic or achromatic), but *where* would that sensation be seen to be?—not, certainly, in the half dozen different places in which the nerve fibre is pinched. The sensations would one and all seem to be in that direction in the external world which is in one-to-one correspondence with the rod or cone in the retina from which the fibre which is pinched has come. We have the same situation when the man who has had his leg amputated at the knee has a nerve fibre pinched—he feels a sensation of tickling (or something), but he feels it not in his knee but in the ball of his foot.[2] It is therefore plain that in the case of the

Blue Arcs of the Retina no possible nerve current evoked in adjoining fibres by the fibre which is conveying the red-light exertation would enable you to "see" those fibres, the actual place-coefficients attached to a sensation produced in this way would be entirely wrong.

There is another insuperable difficulty in resorting to any sort of a nerve-fibre stimulation as the cause of the reddish blue arcs. An after image of these blue arcs can be obtained; one gets a sensation which meets the requirements of the after image both in achromatic intensity and in chroma—it is of a slightly greenish yellow. But an after image does not occur after a stimulation of the visual mechanism by an electric current. This one would have been inclined instinctively to take as self-evident; but it has now been proved beyond question by the work of Lazareff, in which he finally demonstrates that nerve action is not (and cannot be) subject to fatigue.[3]

That nerve fibre when stimulated gives off some sort of an emanation which enables it to take its own photograph has now been proved by Nodan;[4] and it is perfectly easy to suppose that this "emanation" may be of the nature of visible light or at least of a nature to produce visible light by means of fluorescence, which is known to occur in the retina.

Various details of this phenomenon—the blue glow, which has been entirely overlooked by most observers, the absence of the effect in the case of individuals who have myelinated fibres in this region of the retina, and other considerations—make it very certain that its real cause is a bioluminescence on the part of the stimulated nerve fibres. It would follow that all nerves when stimulated shine by their own light—a light which is invisible, of course, when the nerves are not non-myelinated. One has not discovered this before, because, although one devotes much time to studying the stimulated nerve fibre, one does not do it in an absolutely dark room.

[1] This rare work of Purkinje—so rare that Gertz reproduces what Purkinje says about this phenomenon, because so few of his readers will be able to see his book—has now been reproduced in Czecho-Slovakia. Purkinje, Johann Evangelista, *Opera Omnia*. Praze: C. Calve, 1919.

[2] To express this situation the psychologist has need of a new term. It is absurd to apply the phrase "local sign" to the case of the knee and the foot. "Sign" is a wrong term anyway, for there is every reason to think that we have here to do with an actual element of consciousness, and not at all with a *sign* simply. I propose "place-coefficients" as the proper term for this form of *attachment* to the pure sensation.

[3] Lazareff, *Comptes rendus de l'Academie des Sciences*, 1925.

Katz, David: "Die Illusionen der Amputirten," *Beiheft zur Zts. f. angewandten Psychologie*, Leipzig, Barth, 1921.

[4] Nodan, Alfred: *Ibid.*, 1923–1926.

THE METABOLIC EFFECT OF ENEMATA OF ALCOHOL, DEXTROSE AND LEVULOSE IN HUMANS

BY THORNE M. CARPENTER

NUTRITION LABORATORY OF THE CARNEGIE INSTITUTION OF WASHINGTON, BOSTON, MASS.

Communicated April 28, 1926

The usual path of ingestion of food and drugs by man is by way of the mouth. Under some conditions, such as congenital or accidental obstruction of the alimentary canal, unconsciousness, anesthesia or unwillingness of the individual to coöperate (insanity), it is necessary to resort to other paths of ingestion. Those commonly used are the intravenous, subcutaneous, intraperitoneal and rectal injections of the substance needed or under observation. Any study of the metabolic effect of one of these methods of ingestion has therefore a practical significance, as it results in an increased knowledge regarding the effectiveness of the method. Such an investigation has another interest. It is conceivable that when a substance is introduced into the circulation directly or ingested at a different point in the alimentary canal (e.g., the rectum), its path of distribution may be different than when the ingestion is by mouth, and the character and speed of metabolism may also be altered.

This communication has to deal with the effects of enemata of ethyl alcohol, dextrose and levulose upon the metabolism of humans observed in an investigation which began as a part of the program[1] of study of the Nutrition Laboratory on the physiological and psychological effects of ethyl alcohol. Other experiments were added in order to include for comparative purposes the effect of rectal injection of solutions of dextrose and levulose.

The subjects were four medical students presumably in good health. Problems studied were how completely and rapidly the materials were absorbed, what effect, if any, the absorption of these materials had upon the volume and some constituents of the urine, whether the pulse-rate was changed and to what extent the material absorbed was metabolized as indicated by a study of the respiratory exchange. The last was the chief problem of the investigation. The interpretation of the results depended in the main upon the course of the respiratory quotient.

The respiratory quotient is the relation by volume of the carbon dioxide eliminated to the oxygen absorbed when a substance is oxidized. For the utilization of fat the value is 0.713, for carbohydrate 1.00 and for the portion of protein oxidized in the body it is 0.81. The respiratory quotient of alcohol is lower than any of these, namely, 0.667, so that if alcohol is metabolized in the body in such a manner that no other excretory products

are formed than carbon dioxide and water, the result will be a lower respiratory quotient than that existing before alcohol was ingested. That this occurs when alcohol is given by mouth has been shown by a previous investigation in this Laboratory.[2] Oxidation of dextrose or levulose, on the other hand, gives a respiratory quotient of 1.00 and ingestion by mouth results in a rise in the respiratory quotient,[3] the greater and quicker change coming after the taking of levulose.

The absorption of the substances used in the enemata of the present study was determined indirectly; that is to say, after a solution had been retained a given time in the rectum and large intestine, a cleansing enema was given to remove the unabsorbed material, and to be sure that this removal was complete, two and sometimes three such enemata were used. The amounts of the unabsorbed substance thus secured were then determined by suitable chemical methods.

Urine was collected in the period before an experiment (preliminary) and for the time just prior to and following the rectal injection (experimental). The volume, total nitrogen and chlorides were determined for these urines. When alcohol was injected, the amount eliminated by the kidneys was also estimated. In a number of special experiments, the urine was collected in short periods. Determination of the alcohol content of these urines was expected to throw light on the metabolism of alcohol. For comparison purposes there were a few experiments in which similar amounts and concentrations of alcohol were taken orally and the alcohol determined likewise in the urine collected in short periods. The pulse-rate was taken periodically in all experiments in which the respiratory exchange was measured. The latter was determined by means of the gasometer method, in which the expired air is separated from the inspired air by valves and the expired air is collected in a gasometer and analyzed. In several series of experiments a chamber[4] respiration apparatus was used. A 0.6 per cent solution of sodium chloride was used alone in a number of experiments as a control solution to determine the effect of rectal injection of a liquid upon the several physiological and chemical processes.

Ethyl alcohol is very rapidly and nearly completely absorbed when given in enemata. In 31 experiments in which quantities of alcohol varying from 11.0 grams in 220 cc. of 5 per cent solution to 60.8 grams in 810 cc. of 7.5 solution were injected, there were only two cases in which the absorption was not better than 98 per cent, and in these two instances the duration of the periods during which the solution was retained was 1 hour and 4 minutes, and 2 hours and 3 minutes. Alcohol was almost invariably found in the urines collected. The maximum amount eliminated in this manner was only 1.9 per cent of the alcohol given. This occurred in one of the experiments in which urines were collected in short periods and as often as possible, after the giving of 37.5 grams of alcohol in a 5 per cent solu-

tion, the collection lasting about 7 hours. The quantity of urine eliminated after alcohol was given was nearly always increased over that observed in the preliminary period before the experiment. In spite of this, the nitrogen and chloride eliminated were decreased after alcohol, the nitrogen slightly and the chloride to a more marked degree, indicating that the effect of the alcohol was to promote the elimination of water without removing other constituents of the tissues at the same time.

The most striking effect of the rectal injection of alcohol was the marked rise in pulse-rate. Previous researches of Dodge and Benedict,[5] Higgins[2] and Miles[6] have shown that the effect of alcohol is either to prevent the natural fall in pulse-rate, which would take place in the course of an experiment if no alcohol were given, or to produce a slight increase in the pulse-rate. In other words, there was usually a relative acceleration. This amounted to only a few beats at the most. In the experiments with rectal injection, however, the increases in pulse-rate were of much greater magnitude. The most marked effect observed was in a series with a subject who slept in a respiration chamber for the entire night after the rectal injection of 37.5 grams of alcohol in 500 cc. of water. In the five experiments of the series, the pulse-rate rose in the course of $6^1/_2$ hours from an average level of 62 to an average of 73, and in an individual experiment it rose from 66 to 87 in $6^1/_2$ hours. Rectal injection of alcohol has a more specific influence upon the pulse-rate than the giving of like quantities by mouth.

The respiratory exchange indicated very clearly a utilization of alcohol. The respiratory quotient began to lower within two hours after the rectal injection of alcohol and the greater the quantity of alcohol given, the more marked and persistent was this effect on the quotient. In the experiments by Higgins[2] the decrease in the respiratory quotient commenced much sooner. Also in experiments in this research when alcohol was given by mouth, the respiratory quotient was lowered within one-half hour. The studies of the urine, in which the alcohol elimination was determined by periods, indicated that the peak of concentration occurred at about the same time whether ingestion was by mouth or by rectum, that is, within the first hour and one-half, and that the alcohol disappeared from the urine in about the same length of time for equal quantities injected, i.e., 25 grams in about 5 hours, an average disappearance by various ways of 5 grams per hour.

Calculated from the respiratory exchange and by means of Rosemann's formula,[7] the maximum amount theoretically metabolized after rectal injection of alcohol was 4.4 grams per hour when 37.5 grams were given in 500 cc. of water. There is an unexplainable inconsistency in the results of the respiratory exchange measurements and the alcohol in urine determinations in that the respiratory quotient in these experiments re-

mained much depressed at a time after the giving of alcohol when urine analysis would indicate a tendency for the alcohol to disappear.

The solutions of dextrose and levulose in 5, 6 and 10 per cent concentrations were not so rapidly or so completely absorbed as similar solutions of alcohol which were retained for the same length of time. The maximum absorption of dextrose was 90 per cent in the case when 30 grams were given and the material injected was retained about $6^1/_2$ hours. The maximum amount absorbed per hour was 7.8 grams in a period of $2^1/_4$ hours.

Levulose was, in general, absorbed slightly more rapidly than dextrose. The apparent absorptions are higher on the average than those for dextrose, several of them being over 90 per cent with a 5 per cent solution. When 50 grams in solution were retained $4^1/_2$ hours, an average of 10.4 grams per hour were absorbed. Indications from the results are that the rate of absorption of dextrose and levulose is greatest at first, and prolonging the time of retention of the material injected does not increase proportionately the amount of absorption.

In the majority of experiments with dextrose the respiratory quotients were definitely raised, thus indicating an increased utilization of carbohydrate in the metabolism. The maximum increase in carbohydrate utilized calculated from the changes in the respiratory quotient was 4 grams per hour. If all of the carbohydrate utilized after the injection of dextrose was supplied by the dextrose absorbed, the absorption recorded would hardly suffice for more than two or three hours. If one considered only the increase in utilization of carbohydrate, the amount absorbed would be sufficient for a much longer time.

On the other hand, the rectal injection of levulose does not, in the majority of cases, produce any marked rise in the respiratory quotient. This is in direct contrast to its effect when introduced by mouth. When it is introduced by mouth, it has a greater effect and produces a more prompt increase in the respiratory quotient than does an equal quantity of dextrose. Levulose, however, of all of the substances given by rectum, caused the greatest decrease in nitrogen elimination in the urine, indicating a sparing action upon the protein metabolism. This was not the case with dextrose, which had an effect upon the nitrogen in the urine about equal to that due to alcohol. Dextrose produced a greater increase in the volume of urine than did levulose while, on the other hand, dextrose had less effect upon the elimination of sodium chloride. In other words, with dextrose there were indications that the liquid injected and sodium chloride were eliminated while with levulose, the liquid injected and sodium chloride were retained.

The results with alcohol, as well as with dextrose and levulose, indicate that the metabolism of materials introduced by rectum is different than that produced when they are given by mouth. The usual explanations

of the different effect of rectal injection upon metabolism are that the material is absorbed in such a manner that it goes directly into the systemic circulation and avoids the portal circulation and that differences observed in the results are due to the slowness of absorption. The first explanation seems untenable because all of the indications both from this work and from other researches are that the material passes high enough into the large intestine to be absorbed into the branches of the portal circulation and that consequently it must pass through the liver by means of the portal circulation. Experiments by others with small quantities of sugar given by mouth indicate that it was not the small quantity absorbed which produced the effects in question so that the slowness of absorption can hardly explain the results. From the results in this investigation it is believed that the material introduced by rectum passes into the portal circulation and through the liver and then throughout the body. As a hypothesis it is suggested that the liver, as well as the rest of the alimentary tract, under the conditions of the experiments is not in an active state so that the substance is metabolized throughout the body in a manner similar to the utilization of material in muscular work. Dextrose and alcohol are both substances which can be utilized in the body when given in enemata.

A review of the earlier literature, the details of the experiments, a discussion of the practical applications and theoretical calculations, together with a suggested hypothesis, appear in Publication No. 369 of the Carnegie Institution of Washington.

[1] Benedict and Dodge: *Tentative Plan for a Proposed Investigation into the Physiological Action of Ethyl Alcohol in Man. Proposed Correlative Study of the Psychological Effects of Alcohol on Man.* Privately printed, Boston, Massachusetts, **1913.** See, also, Dodge and Benedict, *Carnegie Inst. Wash. Pub.*, No. **232** (1915), pp. 266–275.

[2] Higgins, *Jour. Pharm. and Exper. Therap.*, 1917, **9**, p. 441.

[3] Benedict and Carpenter, *Carnegie Inst. Wash. Pub.*, No. **261**, 1918, pp. 240–246.

[4] Benedict and Tompkins, *Boston Med. and Surg. Jour.*, 1916, **174**, pp. 857, 898 and 939.

[5] Dodge and Benedict, *Carnegie Inst. Wash. Pub.*, No. **232**, 1915, p. 233.

[6] Miles, *Carnegie Inst. Wash. Pub.*, No. **333**, 1924, p. 225.

[7] Rosemann, *Oppenheimer's Handbuch d. Biochem.*, 1911, **4** (1), p. 423.

ERRATA

These PROCEEDINGS, pp. 335 ff., May, 1926.

In equation (1), $\sum_{j=1}^{3n}$ should read $\sum_{j=4}^{3n}$,

in equation (2), $\sum_{j=1}^{3n-b}$ should read $\sum_{j=4}^{3n-b}$,

in equation (3), $\nu^{(b)}$ should read $\nu_j^{(b)}$,

and on page 337 "k the gas constant per gram" should read "k the gas constant per molecule."—E. O. SALANT